

# Multiscale Inference for High-Frequency Data

Adam Sykulski, Sofia C. Olhede and Grigorios A. Pavliotis

## Abstract

This paper proposes a novel multiscale estimator for the integrated volatility of an Itô process, in the presence of market microstructure noise (observation error). The multiscale structure of the observed process is represented frequency-by-frequency and the concept of the multiscale ratio is introduced to quantify the bias in the realized integrated volatility due to the observation error. The multiscale ratio is estimated from a single sample path, and a frequency-by-frequency bias correction procedure is proposed, which simultaneously reduces variance. We extend the method to include correlated observation errors and provide the implied time domain form of the estimation procedure. The new method is implemented to estimate the integrated volatility for the Heston and other models, and the improved performance of our method over existing methods is illustrated by simulation studies.

## Index Terms

Bias correction; market microstructure noise; realized volatility; multiscale inference; Whittle likelihood.

## I. INTRODUCTION

Over the last few decades there has been an explosion of available data in diverse areas such as econometrics, atmosphere/ocean science and molecular biology. It is essential to use this available data when developing and testing mathematical models in physics, finance, biology and other disciplines. It is imperative, therefore, to develop accurate and efficient methods for making statistical inference in a parametric as well as a non-parametric setting.

Many interesting phenomena in the sciences are inherently multiscale in the sense that there is an abundance of characteristic temporal and spatial scales. It is quite often the case that a simplified, coarse-grained model is used to describe the essential features of the problem under investigation. Available data is then used to estimate parameters in this reduced model [12], [17], [18]. This renders the problem of statistical inference quite subtle, since the simplified models that are being used are compatible with the data only at sufficiently large scales. In particular, it is not clear how and if the high frequency data that is available should be used in the statistical inference procedure.

On the other hand in many applications such as econometrics [32] and oceanography [13] the observed data is contaminated by high frequency observation error. Statistical inference for data with a multiscale structure and for data contaminated by high frequency noise share common features. In particular the main difficulty in both problems is that the model that we wish to fit the data to is not compatible with the data at all scales. This is an example of a model misspecification problem [20, p. 192].

Parametric and non-parametric estimation for systems with multiple scales and/or the usage of high frequency data has been studied quite extensively in the last few years for the two different types of models. First, the problem of estimating the integrated stochastic volatility in the presence of high frequency observation noise has been considered by various authors [1], [36]. Similar models and inference problems have also been studied in the context of oceanic transport [13]. It was assumed in [1], [36] that the observed process consists of two parts, an Itô process  $X_t$  (i.e. the solution of an SDE, which is a semimartingale) whose integrated stochastic volatility (quadratic variation  $\langle X_t, X_t \rangle$ ) we want to estimate, and a high frequency noise component  $\varepsilon_{t_j}$

$$Y_{t_j} = X_{t_j} + \varepsilon_{t_j}. \quad (1)$$

$\{Y_{t_j}\}_{j=1}^{N+1}$  are the sampled observations. The additional noise  $\{\varepsilon_{t_j}\}_{j=1}^{N+1}$  was used to model market microstructure.

It was shown for the model of (1) that using high frequency data leads to asymptotically biased estimators. In particular if all available data is used for the estimation of the quadratic variation of  $X_t$  then *the realized integrated*

A. Sykulski and G. A. Pavliotis are with the Department of Mathematics, Imperial College London, South Kensington Campus, SW7 2AZ London, UK (adam.sykulski03@imperial.ac.uk, g.pavliotis@imperial.ac.uk), tel: +44 20 7594 8564, fax +44 20 7594 8517.

S. C. Olhede is with the Department of Statistical Science, University College London, Gower Street, London WC1 E6BT, UK (s.olhede@ucl.ac.uk), tel: +44 20 7679 8321, fax: +44 20 7383 4703.

volatility  $[Y_t, Y_t]$  converges to the aggregated variance of the differenced observation noise. Subsampling is therefore necessary for the accurate estimation of the integrated volatility. An algorithm for estimating the integrated volatility which consists of subsampling at an optimal sampling rate combined with averaging and an appropriate debiasing step was proposed in [1], [36]. Various other estimators were suggested in [11], [15], [32], [36] for processes contaminated by high frequency nuisance structure.

Secondly, parameter estimation for fast/slow systems of SDEs for which a limiting SDE for the slow variable can be rigorously shown to exist was studied in [24]–[26]. In these papers the problem of making inferences for the parameters of the limiting (coarse-grained) SDE for the slow variable from observed data generated by the fast/slow system was examined. It was shown that the maximum likelihood estimator is asymptotically biased. In order to correctly estimate the parameters in the drift and the diffusion coefficient of the coarse-grained model from observations of the slow/fast system using maximum likelihood, subsampling at an appropriate rate is necessary. The subsampling rate depends on the ratio between the characteristic time scales of the fast and slow variables. A similar problem, with no explicit scale separation, was studied in [7].

All of the papers mentioned above propose inference methods in the time domain. Yet, it would seem natural to analyse multiscale and high frequency properties of the data in the frequency domain. Most of the time domain methods can be put in a unified framework as linear filtering techniques, i.e. as a convolution with a linear kernel, of some time-domain quadratic function of the data. The understanding of these methods is enhanced by studying them directly in the frequency domain, as convolutions in time are multiplications in frequency. Fourier domain estimators of the integrated volatility have been proposed for observations devoid of microstructure features, see [2], [14], [21]. Fourier domain estimators have also been used for estimating noisy Itô processes (i.e. processes of the form 1), see [22], [29], [30], based on smoothing the time domain quantities by using only a limited number of frequencies in the reconstruction.

The bias in the realized integrated volatility of the observed process  $Y_{t_j}$  due to the observation noise  $\varepsilon_{t_j}$  can be understood directly in the frequency domain, since the energy associated with each frequency is contaminated by the microstructure noise process. This bias is particularly damaging at high frequencies. In this article we propose a frequency-by-frequency de-biasing procedure to improve the accuracy of the estimation of the integrated volatility. The proposed estimation method can also be viewed in the time domain as smoothing the estimated autocovariance of the increments of the process, but where the implied time domain smoothing kernel is itself estimated from the observed process.

In this paper we will consider a regularly sampled Itô process with additive white noise  $\varepsilon_{t_j}$  superimposed upon it at each observation point  $t_j$ , cf (1). The Itô process satisfies an SDE of the form

$$dX_t = \mu_t dt + \sigma_t dB_t, \quad X_0 = x_0. \quad (2)$$

$B_t$  denotes a standard one dimensional Brownian motion and  $\mu_t, \sigma_t$  are (in general) Itô processes, see for example the Heston model which is studied in Section III. The Brownian motions driving the three Itô processes can be correlated. The observations and the process are related through

$$Y_{t_j} = X_{t_j} + \varepsilon_{t_j}, \quad j = 1, 2, \dots, N+1, \quad t_j := \frac{j-1}{N}T = (j-1)\Delta t. \quad (3)$$

We assume the data is regularly spaced. The length of the path  $T$  is fixed. The additive noise  $\{\varepsilon_{t_j}\}_{j=1}^{N+1}$  is initially taken to be a white noise process with variance  $\sigma_\varepsilon^2$ , and it is assumed to be independent of the noise that drives the Itô process  $X_t$ . Our main objective is to estimate the integrated volatility,  $\langle X, X \rangle_T = \int_0^T \sigma_t^2 dt$  of the Itô process  $\{X_t\}$ , from the set of observations  $\{Y_{t_j}\}_{j=1}^{N+1}$ . In the absence of market microstructure noise (i.e., when  $Y_{t_j} = X_{t_j}$ ,  $j = 1, \dots, N+1$ ) the integrated volatility can be estimated from the realized integrated volatility of the process  $\{Y_t\}$  [32]. In the presence of market microstructure noise this is no longer true, see also [36], and a different estimation procedure is necessary.

The proposed estimator can be described roughly as follows. Let  $\{J_k^{(X)}\}$  denote the Discrete Fourier Transform (DFT) of the differenced sampled  $X_t$  process, and similarly for  $Y_{t_j}$  and  $\varepsilon_{t_j}$ . The integrated volatility can be written in terms of the inverse DFT of the variance of  $J_k^{(X)}$ . We calculate the bias in the variance of  $J_k^{(Y)}$ , when using its sample estimator to estimate the variance of  $J_k^{(X)}$ . The high frequency coefficients are heavily contaminated by the microstructure noise. With a formula for the bias it is possible to debias the estimated variance of the Fourier transform at every frequency, with the unknown parameters of the bias estimated using the Whittle likelihood [34],

[35]. This produces a debiased estimator of the integrated volatility via an aggregation of the estimated variance, and we show also that the variance of the proposed estimator is reduced by the debiasing.

Our estimator shows highly competitive mean square error performance; it also has several advantages over existing estimators. First, it is robust with respect to the signal to noise ratio; furthermore, it is easy to formulate and to implement; in addition, it readily generalizes to the case of correlated observation errors (in time). Finally, the properties of our estimator are transparent using frequency domain analysis.

The rest of the paper is organized as follows. In Section II we introduce our estimator and present some of its properties, stated in Theorems 1 and 2. We also discuss the time-domain understanding of the proposed method and the extension of the method to the case where the observation noise is correlated. In Section III we present the results of Monte Carlo simulations for our estimator. Section IV is reserved for conclusions. Various technical results are included in the appendices.

## II. ESTIMATION METHODS

Let  $\{Y_{t_j}\}$  be given by (3), where the noise  $\{\varepsilon_{t_j}\}$  is independent of  $\{X_{t_j}\}$ , is zero-mean and its variance at any time is equal to  $\sigma_\varepsilon^2$ . The simplest estimator of the integrated volatility of  $X_t$  would ignore the high frequency component of the data and use the realized integrated volatility of the observed process. The realized integrated volatility is given by

$$\widehat{\langle X, X \rangle}_T^{(b)} = [Y, Y]_T \equiv \sum_{j=1}^N (Y_{t_{j+1}} - Y_{t_j})^2 = \mathcal{O}\left(\frac{1}{\Delta t}\right). \quad (4)$$

This estimator is both inconsistent and biased, see [15]. For comparative purposes, we define also the realized integrated volatility of the sampled process  $\{X_{t_j}\}$ :

$$\widehat{\langle X, X \rangle}_T^{(u)} = [X, X]_T \equiv \sum_{j=1}^N (X_{t_{j+1}} - X_{t_j})^2 = \mathcal{O}(N\Delta t) = \mathcal{O}(1). \quad (5)$$

This cannot be used in practice as  $X_{t_j}$  is not directly observed. Both these are estimators of the integrated volatility (quadratic variation) of  $X$ .

### A. Fourier Domain Properties

We shall start by deriving an alternative representation of (4) to motivate further development. Firstly we define the increment process of a sample from a generic time series  $U_{t_j}$ ,  $j = 1, \dots, N+1$  by  $\Delta U_{t_j} = U_{t_{j+1}} - U_{t_j}$ ,  $j = 1, \dots, N$ , and then the discrete Fourier Transform of  $\Delta U_{t_j}$  by  $J_k^{(U)}$  as by [27][p. 206]

$$J_k^{(U)} = \sqrt{\frac{1}{N}} \sum_{j=1}^N \Delta U_{t_j} e^{-2\pi i t_j f_k}, \quad f_k = \frac{k}{T}, \quad U = X, Y, \varepsilon. \quad (6)$$

Our proposed estimator will be based on examining the second order properties of  $\{J_k^{(Y)}\}$ .  $|J_k^{(Y)}|^2$  is the *periodogram* [5] defined for a time series and is an inefficient estimator of  $\text{var}\{J_k^{(Y)}\} = \mathcal{S}_{k,k}^{(X)}$ . Firstly we examine the properties of  $\{J_k^{(X)}\}$ . We have, with  $\bar{\mu}_j = \frac{1}{\Delta t} \int_{(j-1)\Delta t}^{j\Delta t} \mu_s ds$  denoting the local average of  $\mu_t$ ,

$$\begin{aligned} \Delta X_{t_j} &= \int_{(j-1)\Delta t}^{j\Delta t} [\mu_s ds + \sigma_s dW_s] = \bar{\mu}_j \Delta t + \int_{(j-1)\Delta t}^{j\Delta t} \sigma_s dW_s, \\ J_k^{(X)} &= \sqrt{\frac{1}{N}} \sum_{j=1}^N \Delta X_{t_j} e^{-2i\pi \frac{kj}{N}} = \sqrt{\frac{1}{N}} \sum_{j=1}^N \left[ \Delta t \bar{\mu}_j + \int_{(j-1)\Delta t}^{j\Delta t} \sigma_s dW_s \right] e^{-2i\pi \frac{kj}{N}} \\ &= \mathcal{O}\left(\Delta t^{1/2} \frac{T}{k}\right) + \sqrt{\frac{1}{N}} \sum_{j=1}^N \int_{(j-1)\Delta t}^{j\Delta t} \sigma_s dW_s e^{-2i\pi \frac{kj}{N}}. \end{aligned} \quad (7)$$

We define

$$\tilde{J}_k^{(X)} = \sqrt{\frac{1}{N}} \sum_{j=1}^N \int_{(j-1)\Delta t}^{j\Delta t} \sigma_s dW_s e^{-2i\pi \frac{kj}{N}},$$

and this to leading order approximates  $J_k^{(X)}$  as  $\Delta t \rightarrow 0$  for all but a few frequencies. We can also note that, since  $\mu_s$  is an Itô process, it has almost surely continuous paths, which implies that

$$\Delta t^2 \sum_{k=0}^{N-1} \left| \sqrt{\frac{1}{N}} \sum_{j=1}^N \bar{\mu}_j e^{-2i\pi \frac{kj}{N}} \right|^2 \sim \sum_k \left| \frac{1}{\sqrt{N}} \Delta t N/k \right|^2 = \mathcal{O}(\Delta t) \quad (8)$$

$$\Delta t \sum_{k=0}^{N-1} \left| \sqrt{\frac{1}{N}} \sum_{j=1}^N \bar{\mu}_j e^{-2i\pi \frac{kj}{N}} \right| \sim \sum_k \left| \frac{1}{\sqrt{N}} \Delta t N/k \right| = \mathcal{O}(\log(\Delta t) \sqrt{\Delta t}), \quad (9)$$

as  $\Delta t \sum_j \bar{\mu}_j e^{-2i\pi \frac{kj}{N}} = \mathcal{O}(\frac{T}{k})$ . So we only need, to leading order, calculate  $\sum_{k=0}^{N-1} |\tilde{J}_k^{(X)}|^2 = \mathcal{O}(1)$  when calculating the properties of  $\sum_{k=0}^{N-1} |J_k^{(X)}|^2$  from (8) and (9). More formally we note that

$$\sum_{k=0}^{N-1} |J_k^{(X)}|^2 = \sum_{k=0}^{N-1} |\tilde{J}_k^{(X)}|^2 + \mathcal{O}(\log(\Delta t) \sqrt{\Delta t}).$$

We need to determine the first and second order structure of  $\{\tilde{J}_k^{(X)}\}_k$ . In general  $\{\tilde{J}_k^{(X)}\}_k$  is a complex-valued random vector, which may not be a sample from a multivariate Gaussian distribution. The covariance matrix of a complex random vector  $\mathbf{Z}$  is given by  $\text{cov}\{\mathbf{Z}, \mathbf{Z}\} = \mathbb{E}\{\mathbf{Z}\mathbf{Z}^H\} - \mathbb{E}\{\mathbf{Z}\}\mathbb{E}\{\mathbf{Z}\}^H$  [23], [28]. We have

$$\mathbb{E}\{\tilde{J}_k^{(X)}\} = 0, \quad k = 1, \dots, N-1.$$

Furthermore, with  $\tilde{S}_{k_1, k_2}^{(X)} = \mathbb{E}\{\tilde{J}_{k_1}^{(X)} \tilde{J}_{k_2}^{(X)*}\}$ ,

$$\begin{aligned} \tilde{S}_{k_1, k_2}^{(X)} &= \frac{1}{N} \mathbb{E} \left\{ \sum_{n=1}^N \int_{(n-1)\Delta t}^{n\Delta t} \sum_{l=1}^N \int_{(l-1)\Delta t}^{l\Delta t} \sigma_s dW_s \sigma_t dW_t e^{-2i\pi(\frac{k_1 n}{N} - \frac{k_2 l}{N})} \right\} \\ &= \frac{1}{N} \sum_{n=1}^N \int_{(n-1)\Delta t}^{n\Delta t} \sum_{l=1}^N \int_{(l-1)\Delta t}^{l\Delta t} \mathbb{E}\{\sigma_s dW_s \sigma_t dW_t\} e^{-2i\pi(\frac{k_1 n}{N} - \frac{k_2 l}{N})} \\ &= \frac{1}{N} \sum_{n=1}^N \int_{(n-1)\Delta t}^{n\Delta t} \sum_{l=1}^N \int_{(l-1)\Delta t}^{l\Delta t} \mathbb{E}\{\sigma_s \sigma_t\} \delta_{nl} \delta(t-s) e^{-2i\pi(\frac{k_1 n}{N} - \frac{k_2 l}{N})} ds dt. \end{aligned}$$

In particular we have that

$$\begin{aligned} \tilde{S}_{k, k}^{(X)} &= \frac{1}{N} \int_0^T \mathbb{E}\{\sigma_s^2\} ds + \mathcal{O}(\Delta t^2) := \bar{\sigma}_X^2 + \mathcal{O}(\Delta t^2) \\ &= \frac{\langle X, X \rangle_T}{N} + \mathcal{O}(\Delta t^2), \end{aligned} \quad (10)$$

where the error terms are due to the Riemann approximation to an integral and thus it follows that

$$\sum_{k=0}^{N-1} \tilde{S}_{k, k}^{(X)} = \int_0^T \mathbb{E}\{\sigma_s^2\} ds + \mathcal{O}(\Delta t). \quad (11)$$

$\bar{\sigma}_X^2$  does not depend on the value of  $k$  but is constant *irrespective* of the value of  $k$ . Malliavin and Mancino [21] in contrast under very light assumptions show how the Fourier coefficients of  $\{\sigma_t^2\}$  can be calculated from the Fourier coefficients of  $dX_t$ , using a Parseval-Rayleigh relationship, see also [22], [30]. We can from (10) make a stronger link from the Fourier transform to the integrated volatility than that of the Parseval-Rayleigh relationship, and shall use this ‘uniformity of energy’ to estimate the microstructure bias.

We note that the covariance between different frequencies is given by:

$$\begin{aligned}\tilde{\mathcal{S}}_{k_1, k_2}^{(X)} &= \frac{1}{N} \sum_{n=1}^N \int_{(n-1)\Delta t}^{n\Delta t} \mathbb{E} \{ \sigma_t^2 \} dt e^{-2i\pi n(\frac{k_1}{N} - \frac{k_2}{N})} \\ &= \frac{1}{N} \int_0^T \mathbb{E} \{ \sigma_t^2 \} dt e^{-2i\pi t(\frac{k_1}{T} - \frac{k_2}{T})} + \mathcal{O}(\Delta t^2).\end{aligned}$$

Let  $\Sigma(f) := \int_0^T \mathbb{E} \sigma_t^2 e^{-2i\pi f t} dt$ . We can bound the size of  $\Sigma(\frac{k_1}{T} - \frac{k_2}{T})$  as  $|k_1 - k_2|$  increases. As  $\mathbb{E} \{ \sigma_t^2 \}$  is smooth in  $t$  the modulus of the covariance can be bounded for increasing  $|k_1 - k_2|$ , as the Fourier transform  $\Sigma(\frac{k_1}{T} - \frac{k_2}{T})$  decays proportionally to  $|k_1 - k_2|^{-\alpha-1}$  where  $\alpha$  is the number of smooth derivatives of  $\mathbb{E} \{ \sigma_t^2 \}$ . We can also directly note that the variance of the discrete Fourier transform of the noise is precisely (this is *not* a large sample result)

$$\mathcal{S}_{k_1, k_2}^{(\varepsilon)} = \sigma_\varepsilon^2 |2 \sin(\pi f_{k_1} \Delta t)|^2 \delta_{k_1, k_2}, \quad (12)$$

by virtue of being the first difference of white noise (see also [4]). The naive estimator can therefore be rewritten as, with  $\mathcal{S}_{k_1, k_2}^{(Y)} = \text{cov}\{J_{k_1}^{(Y)}, J_{k_2}^{(Y)}\}$ :

$$\widehat{\langle X, X \rangle}_T^{(b)} = \sum_{j=1}^N \Delta Y_{t_j}^2 = \sum_{k=0}^{N-1} \left| J_k^{(Y)} \right|^2, \quad (13a)$$

$$\widehat{\mathcal{S}}_{k, k}^{(Y)} = \left| J_k^{(Y)} \right|^2, \quad (13b)$$

$$\mathbb{E} \left\{ \widehat{\langle X, X \rangle}_T^{(u)} \right\} = \sum_{k=0}^{N-1} \tilde{\mathcal{S}}_{k, k}^{(X)} + \mathcal{O}(\log(\Delta t) \sqrt{\Delta t}) + \mathcal{O}(\Delta t) \quad (13c)$$

$$\equiv \sum_{k=0}^{N-1} \mathcal{S}_{k, k}^{(X)}. \quad (13d)$$

The Parseval-Rayleigh relationship in (13a) is discussed in [22], and is used in [21]. We shall now develop a frequency domain specification of the bias of the naive estimator.

*Lemma 1:* (Frequency Domain Bias of the Naive Estimator) Let  $X_t$  be an Itô process and assume that the covariance of  $J_{k_1}^{(X)}$  and  $J_{k_2}^{(X)}$  to be  $\mathcal{S}_{k_1, k_2}^{(X)}$  with the chosen sampling. Then the naive estimator of the integrated volatility given by (13) has an expectation given by:

$$\begin{aligned}\mathbb{E} \left\{ \widehat{\langle X, X \rangle}_T^{(b)} \right\} &= \sum_{k=0}^{N-1} \left( \tilde{\mathcal{S}}_{k, k}^{(X)} + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2 \right) + \mathcal{O}(\log(\Delta t) \sqrt{\Delta t}) \\ &= \mathbb{E} \left\{ \widehat{\langle X, X \rangle}_T^{(u)} \right\} + \sum_{k=0}^{N-1} \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2 + \mathcal{O}(\log(\Delta t) \sqrt{\Delta t}) \\ &= \mathcal{O}(1) + \mathcal{O}(\Delta t^{-1}) + \mathcal{O}(\log(\Delta t) \sqrt{\Delta t}).\end{aligned} \quad (14)$$

*Proof:* This result follows from the independence of  $\{\varepsilon_t\}$  and  $\{X_t\}$ , combined with (11) and (12). ■

We notice directly from (14) that the relative frequency contribution of  $\Delta X_t$  and  $\varepsilon_t$ , i.e.  $\mathcal{S}_{k, k}^{(X)}$  compared to the noise contribution  $\sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2$  determines the inherent bias of  $\widehat{\langle X, X \rangle}_T^{(b)}$ . Estimator (13) is inconsistent and biased since it is equivalent to estimator (4), and such a procedure would give an unbiased estimator of the integrated volatility only when  $\sigma_\varepsilon^2 = 0$ . When the estimator is expressed in the time domain the microstructure cannot be disentangled from the Itô process. On the other hand in the frequency domain, from the very nature of a multiscale process, the contributions to  $\widehat{\mathcal{S}}_{k, k}^{(Y)}$  can be disentangled.

### B. Multiscale Modelling

To correct the biased estimator we need to correct the usage of the biased estimator of  $\mathcal{S}_{k,k}^{(X)}$ ,  $\widehat{\mathcal{S}}_{k,k}^{(Y)}$ , at each frequency. We therefore define a new shrinkage estimator [33, p. 155] of  $\mathcal{S}_{k,k}^{(X)}$  by

$$\widehat{\mathcal{S}}_{k,k}^{(X)}(L_k) = L_k \widehat{\mathcal{S}}_{k,k}^{(Y)}. \quad (15)$$

$0 \leq L_k \leq 1$  is referred to as the *multiscale ratio* and its optimal form for perfect bias correction is for an arbitrary Itô process given by

$$L_k = \frac{\mathcal{S}_{k,k}^{(X)}}{\mathcal{S}_{k,k}^{(X)} + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2}. \quad (16)$$

This quantity *cannot* be calculated without explicit knowledge of  $\mathcal{S}_{k,k}^{(X)}$  and  $\sigma_\varepsilon^2$ . We can however use (10) to simplify (16) to obtain

$$L_k = \frac{\overline{\sigma}_X^2}{\overline{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2}. \quad (17)$$

For a fixed  $0 \leq L_k \leq 1$

$$\begin{aligned} \mathbb{E} \left\{ \widehat{\mathcal{S}}_{kk}^{(X)}(L_k) \right\} &= L_k \mathbb{E} \left\{ |J_k^{(Y)}|^2 \right\} \\ &= \overline{\sigma}_X^2 + \mathcal{O} \left( \frac{\Delta t^{3/2}}{k} \right), \end{aligned}$$

where the order terms follow from the continuity of  $\mu_s$ . We can define a new estimator for the true  $L_k$  via:

$$\widehat{\langle X, X \rangle}_T^{(m_3)} = \sum_{k=0}^{N-1} \widehat{\mathcal{S}}_{kk}^{(X)}(L_k),$$

where

$$\mathbb{E} \left\{ \widehat{\langle X, X \rangle}_T^{(m_3)} \right\} = \langle X, X \rangle_T + \mathcal{O} \left( \log(\Delta t) \sqrt{\Delta t} \right).$$

Recall that  $\langle X, X \rangle_T = O(T) = O(1)$ . Consequently, to leading order we can remove the bias from the naive estimator if we know the multiscale ratio. We shall now develop a multiscale understanding of the process under observation and use this to construct an estimator for the multiscale ratio.

### C. Estimation of the Multiscale Ratio

We have a two-parameter description on how the energy should be adjusted at each frequency. We only need to determine estimators of  $\sigma = (\overline{\sigma}_X^2, \sigma_\varepsilon^2)$ . We propose to implement the estimation using the Whittle likelihood methods (see [3] or [34], [35]). For a time-domain sample  $\Delta \mathbf{Y} = (\Delta Y_{t_1}, \dots, \Delta Y_{t_N})$  that is *stationary*, if suitable conditions are satisfied, see for example [8], then the Whittle likelihood approximates the time domain likelihood, with improving approximation as the sample size increases. It is possible to show a number of suitable properties of estimators based on the Whittle likelihood, see [34], [35]. For processes that are not stationary, such conditions are in general not met, and so the function can be used as an objective function to construct estimators, but not as a true likelihood. The Whittle log-likelihood is defined [34], [35] by

$$\begin{aligned} l(\mathcal{S}) &\equiv \log \left[ \prod_{k=1}^{N/2-1} \frac{1}{\mathcal{S}_{kk}^{(Y)}} e^{-\frac{\widehat{\mathcal{S}}_{kk}^{(Y)}}{\mathcal{S}_{kk}^{(Y)}}} \right] \\ &= - \sum_{k=1}^{N/2-1} \log \left( \mathcal{S}_{kk}^{(X)} + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2 \right) - \sum_{k=1}^{N/2-1} \frac{\widehat{\mathcal{S}}_{kk}^{(Y)}}{\mathcal{S}_{kk}^{(X)} + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2}. \end{aligned}$$

If  $\{\Delta X_t\}$  is not stationary, then as long as the total contributions of the covariance of the incremental process can be bounded, using this likelihood will asymptotically (in  $\Delta t^{-1}$ ) produce suitable estimators, as we shall discuss further.

**Definition 2.1:** (Multiscale Energy Likelihood)

The multiscale energy log-likelihood is then defined using (10) as:

$$l(\sigma) = - \sum_{k=1}^{N/2-1} \log \left( \bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2 \right) - \sum_{k=1}^{N/2-1} \frac{\widehat{\mathcal{S}}_{kk}^{(Y)}}{\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2}. \quad (18)$$

We stress that strictly speaking this may not be a (log-)likelihood, but merely a device for determining the multiscale ratio. We maximise this function in  $\sigma$  to obtain a set of estimators  $\widehat{\sigma}$ .

**Theorem 1:** (The Estimated Multiscale Ratio)

The estimated multiscale ratio is given by

$$\widehat{L}_k = \frac{\widehat{\sigma}_X^2}{\widehat{\sigma}_X^2 + \widehat{\sigma}_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2}, \quad (19)$$

where  $\widehat{\sigma}_X^2$  and  $\widehat{\sigma}_\varepsilon^2$  maximise  $\ell(\sigma)$  given in (18).  $\widehat{L}_k$  satisfies

$$\frac{\widehat{L}_k}{L_k} = 1 + \mathcal{O}(\Delta t^{1/4}). \quad (20)$$

*Proof:* See Appendix A. ■

Combining (15) with (19) the proposed estimator of the spectral density of  $\{\Delta X_t\}$  is:

$$\widehat{\mathcal{S}}_{kk}^{(X)}(\widehat{L}_k) = \widehat{L}_k \widehat{\mathcal{S}}_{kk}^{(Y)}, \quad (21)$$

where  $\widehat{L}_k$  is given by (19).

**Theorem 2:** (The Multiscale Estimator of the Integrated Volatility)

Assume that  $\Delta X_{t_j}$  satisfies the conditions of Lemma 1 and Theorem 1. The multiscale estimator of the integrated volatility defined by

$$\widehat{\langle X, X \rangle}_T^{(m_1)} = \sum_{k=0}^{N-1} \widehat{\mathcal{S}}_{kk}^{(X)}(\widehat{L}_k), \quad (22)$$

where  $\widehat{\mathcal{S}}_{kk}^{(X)}(\widehat{L}_k)$  is defined by (21) has a mean and variance given by:

$$\begin{aligned} \mathbb{E} \left\{ \widehat{\langle X, X \rangle}_T^{(m_1)} \right\} &= \sum_{k=0}^{N-1} \mathcal{S}_{kk}^{(X)} + \mathcal{O}(\log(\Delta t) \sqrt{\Delta t}) + \mathcal{O}(\Delta t^{1/4}) \\ &= \int_0^T \mathbb{E} \{ \sigma_t^2 \} + \mathcal{O}(\log(\Delta t) \sqrt{\Delta t}) + \mathcal{O}(\Delta t^{1/4}) \end{aligned}$$

and

$$\text{var} \left\{ \widehat{\langle X, X \rangle}_T^{(m_1)} \right\} = \sum_{k=0}^{N-1} L_k^2 \left| \mathcal{S}_{k,k}^{(Y)} \right|^2 + \mathcal{O}(\Delta t^{1/2}) = \mathcal{O}(\Delta t^{1/2}).$$

*Proof:* See Appendix B. ■

We also note that

$$\begin{aligned} \text{var} \left\{ \widehat{\langle X, X \rangle}_T^{(m_1)} \right\} &= \sum_{k=0}^{N-1} L_k^2 \left| \mathcal{S}_{k,k}^{(Y)} \right|^2 + \mathcal{O}(\Delta t^{1/2}) \\ &< \mathcal{O} \left( \frac{1}{\Delta t} \right) = \text{var} \left\{ \widehat{\langle X, X \rangle}_T^{(b)} \right\}, \end{aligned} \quad (23)$$

unless  $\sigma_\varepsilon = 0$ . We note that the multiscale estimator has *lower* variance than the naive method of moments estimator  $\widehat{\langle X, X \rangle}_T^{(b)}$  due to the fact that  $0 \leq L_k \leq 1$ . We have thus removed bias *and* simultaneously decreased the variance, the latter effect usually being the main purpose of shrinkage estimators. Note that if we knew the true multiscale ratio  $L_k$  and used it rather than  $\widehat{L}_k$  (i.e. used  $\widehat{\langle X, X \rangle}_T^{(m_3)}$ ) then we would expect an estimator from this quantity to recover the same variance as the estimator based on the noise-free observations. This loss of efficiency is inevitable,

as we have to estimate  $L_k$ . Finally we can also construct a Whittle estimator for the integrated volatility by starting from (10) and taking

$$\widehat{\langle X, X \rangle}_T^{(w)} = N \hat{\sigma}_X^2. \quad (24)$$

The sampling properties of  $\widehat{\langle X, X \rangle}_T^{(w)}$  are found in Appendix A, and  $\hat{\sigma}_X^2$  is asymptotically unbiased. The results in Appendix A imply that

$$\text{var} \left\{ \widehat{\langle X, X \rangle}_T^{(w)} \right\} = T \frac{\sigma_\varepsilon}{\tau_X^{1/2}} 16 \tau_X^2 \sqrt{\Delta t}. \quad (25)$$

We see that the variance depends on the length of the time course, the inverse of the signal to noise ratio, the square root of the sampling period and the fourth power of the “average standard deviation” of the  $X_t$  process. We may compare the variance of (23) with the variance of (25), to determine which estimator of  $\widehat{\langle X, X \rangle}_T^{(w)}$  and  $\widehat{\langle X, X \rangle}_T^{(m_1)}$  is preferable. We shall return to this question of relative performance in the examples section, but intuitively argue that  $\widehat{\langle X, X \rangle}_T^{(w)}$  and  $\widehat{\langle X, X \rangle}_T^{(m_1)}$  are more or less the same estimator, with the latter estimator being more intuitive to explain.

#### D. Time Domain Understanding of the Method

We may write the frequency domain estimator of the spectral density of  $\Delta X_t$  in the time domain to clarify some of its properties. We define

$$\hat{s}_\tau^{(X)} = \frac{1}{N} \sum_{k=0}^{N-1} \hat{\mathcal{S}}_{kk}^{(X)} e^{2i\pi \frac{k\tau}{N}}, \quad \tau \in \mathbb{N},$$

which when  $\Delta X_t$  is a stationary process corresponds to the estimated autocovariance sequence of  $\Delta X_t$  using the method of moments estimator [5, Ch. 5]. We then have

$$\hat{\mathcal{S}}_{kk}^{(X)} = L_k \hat{\mathcal{S}}_{kk}^{(Y)}, \quad \hat{s}_\tau^{(X)} = \sum_u \ell_{\tau-u} \hat{s}_u^{(Y)}, \quad (26)$$

and so the estimated autocovariance of the  $\Delta X_t$  process, namely  $\hat{s}_\tau^{(X)}$ , is a smoothed version of  $\hat{s}_\tau^{(Y)}$ . We can therefore view  $\hat{\mathcal{S}}_{kk}^{(X)}$  as the Fourier transform of a smoothed version of the autocovariance sequence of  $\Delta Y_t$ . We let

$$L(f) = \frac{\bar{\sigma}_X^2}{\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f \Delta t)|^2}, \quad (27)$$

be the continuous analogue of  $L_k$ . To find the smoothing kernel we are using we need to calculate

$$\begin{aligned} \ell_\tau &= \frac{1}{N} \sum_{k=0}^{N-1} L_k e^{2i\pi \frac{k\tau}{N}} \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\bar{\sigma}_X^2}{\bar{\sigma}_X^2 + 4\sigma_\varepsilon^2 \sin^2(\pi f)} e^{2i\pi f \tau} df + \mathcal{O}(\Delta t). \end{aligned} \quad (28)$$

Thus utilizing integration in the complex plane (see Appendix C) we obtain that

$$\ell_\tau = \begin{cases} \left( \frac{\sigma_\varepsilon^2}{\bar{\sigma}_X^2} \right)^\tau + \mathcal{O} \left( \left( \frac{\sigma_\varepsilon}{\bar{\sigma}_X} \right)^{2\tau+2} \right) & \text{if } \sigma_\varepsilon^2 < \bar{\sigma}_X^2 \\ \frac{\bar{\sigma}_X}{2\sigma_\varepsilon} \left( 1 - \frac{\bar{\sigma}_X}{\sigma_\varepsilon} \right)^\tau + \mathcal{O} \left( \frac{\bar{\sigma}_X^2}{2\sigma_\varepsilon^2} \left( 1 - \frac{\bar{\sigma}_X}{\sigma_\varepsilon} \right)^\tau \right) & \text{if } \sigma_\varepsilon^2 > \bar{\sigma}_X^2 \end{cases} \quad (29)$$

These are both decreasing sequences in  $\tau$ . We write  $r_\tau = \frac{\bar{\sigma}_X}{2\sigma_\varepsilon} \left( 1 - \frac{\bar{\sigma}_X}{\sigma_\varepsilon} \right)^\tau$ . If we can additionally assume that  $L(f)$  decreases sufficiently rapidly to be near zero by  $f = \frac{1}{\pi}$  then we find that

$$\ell_\tau \approx q_\tau = \frac{\bar{\sigma}_X}{2\sigma_\varepsilon} e^{-\frac{\bar{\sigma}_X}{\sigma_\varepsilon} |\tau|}.$$



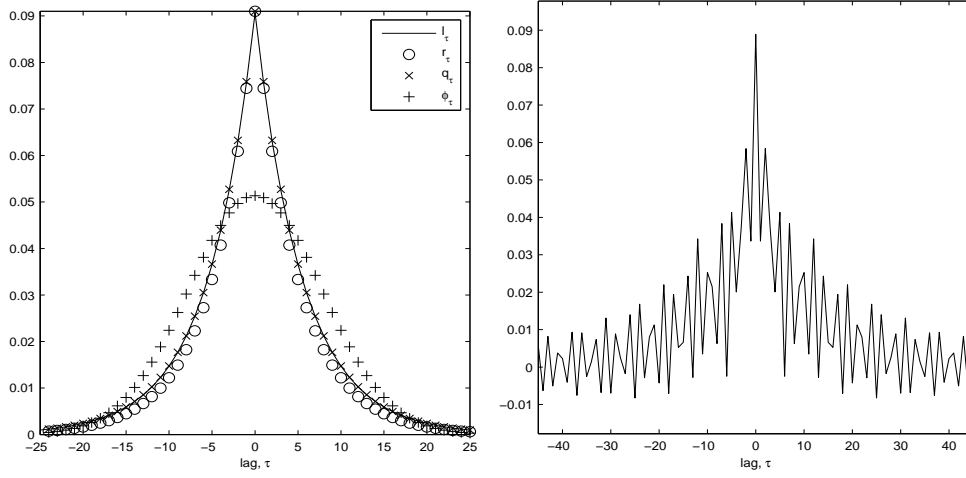


Fig. 1.  $\ell_\tau$  as well as  $r_\tau$  and  $q_\tau$  for a chosen value of the SNR (left). The approximate weighting functions perfectly mirror the exact calculation. We overlay a Gaussian kernel with the same spread for comparison.  $\ell_\tau$  estimated for the MA(6) case (right).

In the limit of no observation noise ( $\frac{\bar{\sigma}_X}{\sigma_\varepsilon} \rightarrow \infty$ ) then this sequence becomes a delta function centered at  $\tau = 0$ . Let us plot these functions, i.e.  $\ell_\tau$ ,  $r_\tau$  and  $q_\tau$  for a chosen case of  $\bar{\sigma}_X^2/\sigma_\varepsilon^2 \approx 0.0331$  (the approximate SNR used in a later example), in Figure 1 (left). We see that theory coincides very well with practise, and almost perfect agreement between the three functions.  $\ell_\tau$  is however a strange choice of kernel, if dictated by the statistical inference problem: it has heavier tails than the common choice of the Gaussian kernel, and is extremely peaked around zero (a Gaussian kernel with the same variance has been overlaid in Figure 1). This is not strange, as we are trying to filter out correlations due to non-Itô behaviour, but counter to our intuition about suitable kernel functions, as the differenced Itô process exhibits very little covariance at any lag but zero, the sharp peak at zero is necessary.

### E. Correlated Errors

In many applications we need to consider correlated observation noise. We assume that despite being dependent the  $\varepsilon_{t_j}$  is a stationary time series. Stationary processes can be conveniently represented in terms of aggregations of uncorrelated white noise processes, using the Wold decomposition theorem [6][p. 187]. We may therefore write the zero-mean observation  $\varepsilon_{t_j}$  as

$$\varepsilon_{t_j} = \sum_{k=0}^{\infty} \theta_{t_k} \eta_{t_j - t_k}, \quad (30)$$

where  $\theta_{t_0} \equiv 1$ ,  $\sum_j \theta_{t_j}^2 < \infty$ , and  $\{\eta_{t_n}\}$  satisfies  $E\{\eta_{t_n}\} = 0$  and  $E\{\eta_{t_n} \eta_{t_m}\} = \sigma_\eta^2 \delta_{n,m}$ , a model also used in [31]. Common practise would involve approximating the variable by a finite number of elements in the sum, and thus we truncate (30) to some  $q \in \mathbb{Z}$ . We therefore model the noise as a Moving Average (MA) process specified by

$$\varepsilon_{t_j} = \eta_{t_j} + \sum_{k=1}^q \theta_{t_k} \eta_{t_j - k}, \quad (31)$$

and the covariance of the DFT of the differenced  $\varepsilon_{t_j}$  process takes the form:

$$\mathcal{S}_{k,k}^{(\varepsilon)} = \sigma_\eta^2 \left| 1 + \sum_{k=1}^q \theta_k e^{2i\pi f k} \right|^2 |2 \sin(\pi f \Delta t)|^2. \quad (32)$$

This leads to defining a new multiscale ratio replacing  $\sigma_\varepsilon^2 |2 \sin(\pi f \Delta t)|^2$  of (17) with  $\sigma_\eta^2 \left| 1 + \sum_{k=1}^q \theta_k e^{2i\pi f k} \right|^2 |2 \sin(\pi f \Delta t)|^2$ . We then obtain a new estimator of  $\mathcal{S}_{kk}^{(X)}$ . In general the value of  $q$  is not known. To simultaneously implement model choice, we need to penalize the likelihood. We define the corrected Aikake information criterion (AICC) by [6, p. 303] (refer to (18) for  $l(\sigma, \theta)$  with  $\sigma_\varepsilon^2 |2 \sin(\pi f \Delta t)|^2$  replaced by  $\sigma_\eta^2 \left| 1 + \sum_{k=1}^q \theta_k e^{2i\pi f k} \right|^2 |2 \sin(\pi f \Delta t)|^2$ )

$$\text{AICC}(\theta) = -2l(\sigma, \theta) + 2 \frac{(p+2)n}{n-p-3}. \quad (33)$$

By minimizing this function, in  $\sigma$ ,  $\theta$  and  $q$ , we obtain the best fitting model for the noise accounting for overfitting by using the penalty term. With this method we retrieve a new multiplier that is applied in the Fourier domain, which corresponds to a new smoother in the Fourier domain, where the smoothing window (and its smoothing width) have been automatically chosen by the data. See an example of such a smoothing window  $\ell_\tau$  in Figure 1 (right). Here  $L_k$  has been estimated from an Itô process immersed in an MA noise process. The spectrum of the MA has a trough at frequency 0.42. We therefore expect to reinforce oscillations at period  $1/0.42 \approx 2.5$ , which is evident from the oscillations of the estimated kernel. For more details of this process see section III-E.

### III. MONTE CARLO STUDIES

In this section we demonstrate the performance of the multiscale estimator through Monte Carlo simulations. We first describe the de-biasing procedure of the estimator for the Heston Model using Fourier domain graphs. We then present bias, variance and mean square error results of various estimators (including the multiscale estimator, the naive estimator and the first-best estimator developed in [36]), for the Heston Model as well as Brownian and Ornstein Uhlenbeck processes. We then consider the case where the sample path in a Heston Model is much shorter and another case where the microstructure noise is greatly reduced. Finally, we consider the case of correlated errors and show how a stationary noise process can be captured using model choice methods and then the integrated volatility can be estimated using the adjusted multiscale estimator.

#### A. The Heston Model

The Heston model is specified in [16]:

$$dX_t = (\mu - \nu_t/2) dt + \sigma_t dB_t, \quad d\nu_t = \kappa(\alpha - \nu_t) dt + \gamma \nu_t^{1/2} dW_t, \quad (34)$$

where  $\nu_t = \sigma_t^2$ , and  $B_t$  and  $W_t$  are correlated 1-D Brownian motions. We will use the same parameter values to the ones that were used in [36], namely  $\mu = .05$ ,  $\kappa = 5$ ,  $\alpha = .04$ ,  $\gamma = .5$  and the correlation coefficient between the two Brownian motions B and W is  $\rho = -.5$ . We set  $X_0 = 0$  and  $\nu_0 = 0.04$ , which is the long time limit of the expectation of the process  $\nu_t$ .<sup>1</sup>

We calculate  $\widehat{\mathcal{S}}_{kk}^{(X)}$  and  $\widehat{\mathcal{S}}_{kk}^{(\varepsilon)}$  directly from simulated data and average across realizations, producing Figure 2, where  $k$  is indicated by its frequency  $f_k = k/N$ , and only plotted for  $k = 0, \dots, N/2 - 1$ , as the spectrum (or  $\mathcal{S}_{kk}^{(X)}$ ) is symmetric. We see directly from these plots that (on average as we showed)  $\widehat{\mathcal{S}}_{kk}^{(X)}$  is constant whilst  $\widehat{\mathcal{S}}_{kk}^{(\varepsilon)}$  is strongly increasing with  $k$ , completely dwarfing the other spectrum at large  $k$ . (11) implies that an equal weighting is given to all frequencies for the differenced Itô process. The noise process will in contrast have a spectrum that is far from flat, and a suitable bias correction would shrink the estimator of  $\mathcal{S}_{kk}^{(X)}$  at higher frequencies.

We also calculate  $\widehat{\mathcal{S}}_{kk}^{(X)}$  and  $\widehat{\mathcal{S}}_{kk}^{(\varepsilon)}$  for one simulated path, displayed in Figure 3. Here we have used the same sample length  $T$  and noise intensity  $\sigma_\varepsilon^2$  as in [36]:  $T = 1$  day and  $\sigma_\varepsilon^2 = 0.0005^2$ . The length of the sample path,  $T = 1$  day or 23,400s with  $\Delta t = 1s$ , corresponds to one trading day, since we take one trading day to be 6.5h long. Notice the different shape of the two periodograms.  $\widehat{\mathcal{S}}_{kk}^{(Y)}$  will not be distinguishable from  $\widehat{\mathcal{S}}_{kk}^{(\varepsilon)}$  at higher frequencies, despite the moderate to low intensity of the market microstructure noise. If we observed the two components  $X_t$  and  $\varepsilon_t$  separately, then the multiscale ratio  $L_k$  could be estimated from  $\widehat{\mathcal{S}}_{kk}^{(X)}$  and  $\widehat{\mathcal{S}}_{kk}^{(\varepsilon)}$  using the method of moments formula. In this case, we would estimate  $L_k$  by the sample Fourier Transform variances

$$\widetilde{L}_k = \frac{\widehat{\mathcal{S}}_{kk}^{(X)}}{\widehat{\mathcal{S}}_{kk}^{(X)} + \widehat{\mathcal{S}}_{kk}^{(\varepsilon)}}. \quad (35)$$

The corresponding estimator of the integrated volatility becomes

$$\widehat{\langle X, X \rangle}_T^{(m_2)} = \sum_{k=0}^{N-1} \widetilde{L}_k \widehat{\mathcal{S}}_{kk}^{(Y)}. \quad (36)$$

The estimated multiscale ratio  $\widetilde{L}_k$ , for the Heston model with the specified parameters, is plotted in Figure 4.

<sup>1</sup> $\lim_{t \rightarrow +\infty} \mathbb{E}\nu_t = \alpha$ .

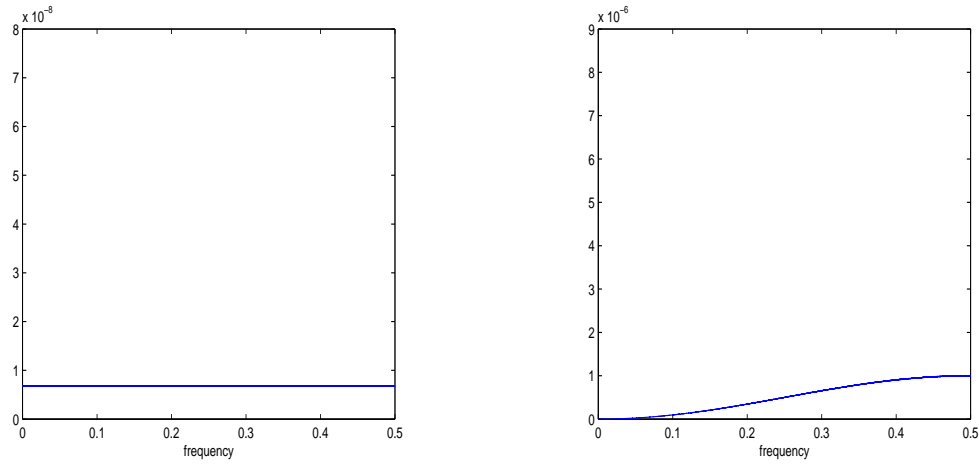


Fig. 2.  $\widehat{\mathcal{S}}_{kk}^{(X)}$  (left) and  $\widehat{\mathcal{S}}_{kk}^{(\varepsilon)}$  (right) averaged over 100,000 realizations. Note the different scaling of the  $y$  axis in the two figures.

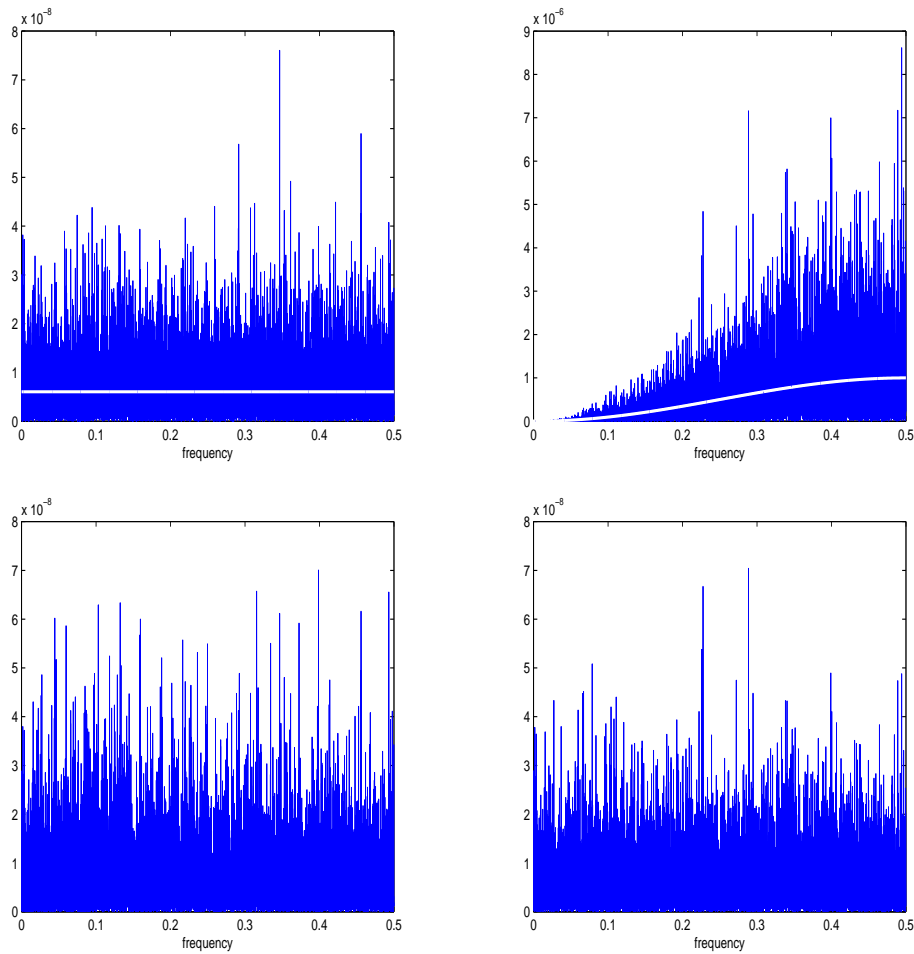


Fig. 3. A realisation of  $\widehat{\mathcal{S}}_{kk}^{(X)}$  (top left), a realisation of  $\widehat{\mathcal{S}}_{kk}^{(\varepsilon)}$  (top right) with the Whittle estimates superimposed and of two biased corrected estimators of  $\mathcal{S}_{kk}^{(X)}$ , using  $\widehat{L}_k \widehat{\mathcal{S}}_{kk}^{(X)}$  (bottom left) and  $\widehat{L}_k \widehat{\mathcal{S}}_{kk}^{(Y)}$  (bottom right). Notice the different scales in the four figures. Estimated spectra are plotted on a linear scale for ease of comparison to the effect of applying  $\widehat{L}_k$ .

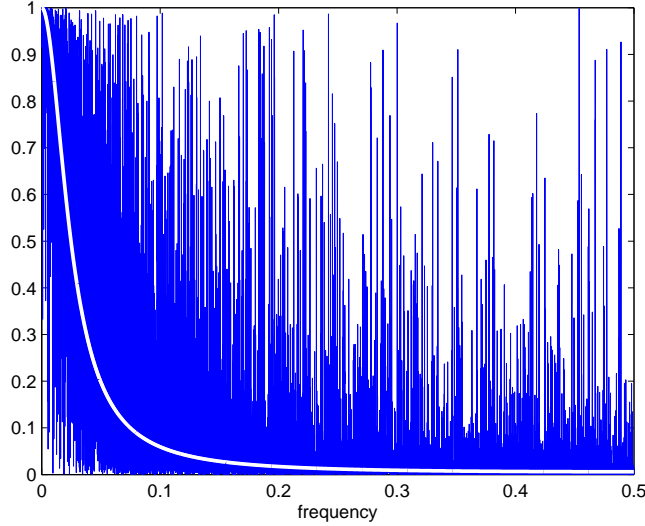


Fig. 4. The method of moments estimate  $\tilde{L}_k$  from a single realisation, with the Whittle estimate (white line) of  $L_k$  superimposed.

The multiscale ratio cannot be estimated using the method of moments in realistic scenarios, as we only observe the aggregated process  $Y_t$  and not the two processes  $X_t$  and  $\varepsilon_t$  separately. Figure 3 displays the estimated multiscale ratio  $\tilde{L}_k$  applied to  $\hat{S}_{kk}^{(Y)}$  over one path realisation. This plot suggests that the energy over the high frequencies has been shrunk and that  $\tilde{L}_k \hat{S}_{kk}^{(Y)}$  is a good approximation to  $\hat{S}_{kk}^{(X)}$ . It therefore seems not unreasonable that the summation of this function across frequencies should make a good approximation to the integrated volatility.

The parameters ( $\hat{\sigma}_X^2$  and  $\hat{\sigma}_\varepsilon^2$ ) are found separately for each path using the MATLAB function `fmincon` on (18). Figure 3 shows  $\hat{\sigma}_X^2$  and  $\hat{\sigma}_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2$  (in white) plotted over the periodograms  $\hat{S}_{kk}^{(X)}$  and  $\hat{S}_{kk}^{(\varepsilon)}$  for one simulated path. The approximated values of  $\bar{\sigma}_X^2$  and  $\sigma_\varepsilon^2$  are quite similar to the averaged periodograms of Figure 2; in fact the accuracy of the new estimator depends on how consistently these parameters are estimated in the presence of limited information from the sampled process  $Y_t$ . Figure 4 shows the corresponding estimated multiscale ratio  $\hat{L}_k$  (in white) from this simulated path, as defined in (19). The function decays, as expected, so that it will remove the high-frequency microstructure noise in the spectrum of  $Y_t$ ; the ratio is also a good approximation of  $\tilde{L}_k$ . Figure 3 shows  $\hat{L}_k \hat{S}_{kk}^{(Y)}$ , which is again similar to  $\hat{S}_{kk}^{(X)}$ . It would appear that the new estimator has successfully removed the microstructure effect from each frequency.

It is worth noting that the ratios  $L_k$  and  $\hat{L}_k$  quantify the effect of the multiscale structure of the process. If  $\sigma_\varepsilon^2$  is zero (ie. there is no microstructure noise), then no correction will be made to the spectral density function (the ratio will equal 1 at all frequencies). So in the case of zero microstructure noise, the estimate would recover  $\hat{S}_{kk}^{(X)}$  and from (13) the estimate of the integrated volatility would simply be the realized integrated volatility of the observable process.

We investigate the performance of the multiscale estimator using Monte Carlo simulations. In this study 50,000 simulated paths are generated. Table I displays the results of our simulation, where biases, variances and errors are calculated using a Riemann sum approximation of the integral

$$\frac{T}{N} \sum_{i=1}^N \sigma_i^2 = \int_0^T \sigma_t^2 dt. \quad (37)$$

The two estimators  $\widehat{\langle X, X \rangle}_T^{(u)}$  and  $\widehat{\langle X, X \rangle}_T^{(m_2)}$  (see (5) and (36) respectively) are both included for comparison, even though these require use of the unobservable  $X_t$  process. The performance of the first-best estimator in [36] (denoted by  $\widehat{\langle X, X \rangle}_T^{(s_1)}$ ) is also included as a well-performing and tested estimator using only the  $Y_t$  process, as is the naive estimator of the realized volatility on  $Y_t$  at the highest frequency,  $\widehat{\langle X, X \rangle}_T^{(b)}$ , given in (4) (the fifth-best

	Sample bias	Sample variance	Sample RMSE
$\widehat{\langle X, X \rangle}_T^{(b)}$	$1.17 \times 10^{-2}$	$1.80 \times 10^{-8}$	$1.17 \times 10^{-2}$
$\widehat{\langle X, X \rangle}_T^{(s_1)}$	$6.44 \times 10^{-7}$	$2.76 \times 10^{-10}$	$1.66 \times 10^{-5}$
$\widehat{\langle X, X \rangle}_T^{(m_1)}$	$2.90 \times 10^{-7}$	$2.59 \times 10^{-10}$	$1.61 \times 10^{-5}$
$\widehat{\langle X, X \rangle}_T^{(w)}$	$2.63 \times 10^{-7}$	$2.59 \times 10^{-10}$	$1.61 \times 10^{-5}$
$\widehat{\langle X, X \rangle}_T^{(m_2)}$	$1.39 \times 10^{-8}$	$2.07 \times 10^{-10}$	$1.44 \times 10^{-5}$
$\widehat{\langle X, X \rangle}_T^{(u)}$	$1.20 \times 10^{-8}$	$2.06 \times 10^{-10}$	$1.44 \times 10^{-5}$

TABLE I  
SIMULATION STUDY COMPARING THE NEW ESTIMATOR WITH THE BEST ESTIMATOR OF [36].

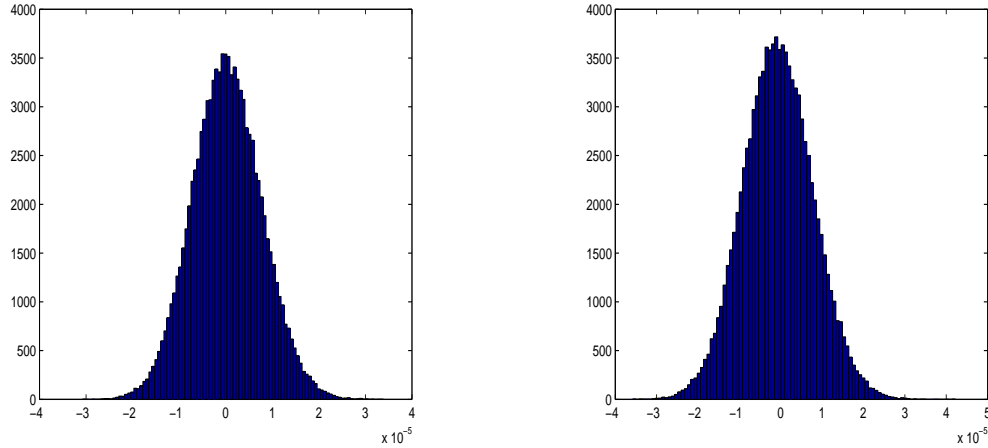


Fig. 5. The histograms of the observed bias of the proposed estimator (a), and the first-best estimator (b), over 100,000 sample paths.

estimator in [36]). We also include the performance of  $\widehat{\langle X, X \rangle}_T^{(w)}$ , defined in (24).

Table I shows that the new estimator,  $\widehat{\langle X, X \rangle}_T^{(m_1)}$ , is competitive with the first-best approach in [36],  $\widehat{\langle X, X \rangle}_T^{(s_1)}$ , as an estimator of the integrated volatility for the Heston model with the stated parameters. For this simulation the new method performed marginally better. The similar performance of the two estimators is quite remarkable, given their different approach; both estimators involve a bias-correction, [36] perform this globally by weighting different sampling frequencies, whilst we correct locally at each frequency. The realized integrated volatility of  $Y_t$  at the highest frequency,  $\widehat{\langle X, X \rangle}_T^{(b)}$ , produces disastrous results, as expected.

We also note that  $\widehat{\langle X, X \rangle}_T^{(m_1)}$  performs more or less identically to  $\widehat{\langle X, X \rangle}_T^{(w)}$ . These two estimators can almost be used interchangeably due to the invariance property of a maximum likelihood estimator. This observation is born out by our simulation studies, and we henceforth only report results for  $\widehat{\langle X, X \rangle}_T^{(m_1)}$ . Note that the variance of  $\widehat{\langle X, X \rangle}_T^{(w)}$  can be found from (25). To compare theory with simulations we note that the average estimated standard deviation is  $1.6093 \times 10^{-5}$  whilst the expression for the variance to leading order gives an expression for the standard deviation of  $\left[ \text{var} \left\{ \widehat{\langle X, X \rangle}_T^{(w)} \right\} \right]^{1/2} = 1.0246 \times 10^{-5}$ , using the parameter values of  $\bar{\sigma}_X^2 \approx 6.8 \times 10^{-9}$  and  $\sigma_\varepsilon^2 \approx 2.5 \times 10^{-7}$ .

A histogram of the observed bias of the new estimator is plotted in Figure 5 along with a histogram of the observed bias of the first-best estimator in [36]. The observed bias of our estimator follows a Gaussian distribution centred at zero, suggesting that this estimator is unbiased, as our results claim to be true. Comparing our estimator to the first-best estimator, it can be seen that the new estimator has similar magnitudes of error also (hence the similar Root Mean Square Error (RMSE)).

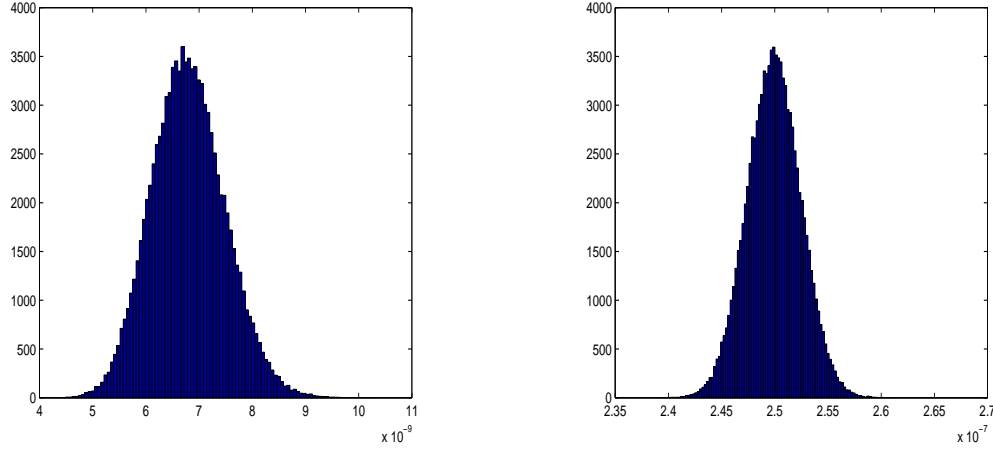


Fig. 6. The histograms of the estimated  $\hat{\sigma}_X^2$  (a) and  $\hat{\sigma}_\epsilon^2$  (b).

	Sample bias	Sample variance	Sample RMSE
$\widehat{\langle X, X \rangle}_T^{(b)}$	$1.17 \times 10^{-2}$	$1.77 \times 10^{-8}$	$1.17 \times 10^{-2}$
$\widehat{\langle X, X \rangle}_T^{(s_1)}$	$6.52 \times 10^{-7}$	$2.68 \times 10^{-11}$	$5.22 \times 10^{-6}$
$\widehat{\langle X, X \rangle}_T^{(m_1)}$	$3.02 \times 10^{-7}$	$1.98 \times 10^{-11}$	$4.46 \times 10^{-6}$
$\widehat{\langle X, X \rangle}_T^{(m_2)}$	$1.96 \times 10^{-9}$	$6.93 \times 10^{-13}$	$8.32 \times 10^{-7}$
$\widehat{\langle X, X \rangle}_T^{(u)}$	$3.79 \times 10^{-9}$	$5.44 \times 10^{-13}$	$7.38 \times 10^{-7}$

TABLE II  
SIMULATION STUDY FOR THE BROWNIAN PROCESS.

The new estimator requires calculation of  $\hat{\sigma}_X^2$  and  $\hat{\sigma}_\epsilon^2$  which will vary over each process due to the limited information given from the  $Y_t$  process. The stability of this estimation is of great importance if the estimator is to perform well. Figure 6 shows the distribution of the parameters  $\hat{\sigma}_X^2$  and  $\hat{\sigma}_\epsilon^2$  over the simulated paths. The parameter estimation is quite consistent, with all values estimated within a narrow range. Figure 2 suggests that these estimates are roughly unbiased; as  $\bar{\sigma}_X^2 \approx 6.8 \times 10^{-9}$  and  $\sigma_\epsilon^2 \approx 2.5 \times 10^{-7}$  (as  $\sigma_\epsilon^2 |2 \sin(\pi f_k)|^2 \approx 1 \times 10^{-6}$ , at  $f_k = 0.5$ ).

### B. Brownian Process and Ornstein Uhlenbeck Process

We repeated our simulations for a Brownian Process given by:

$$dX_t = \sqrt{2\sigma_t^2} dB_t, \quad (38)$$

where  $\sigma_t^2 = 0.01$ . We otherwise keep the same simulation setup as before with 50,000 simulated paths of length 23,400. The results are displayed in Table II. The new estimator,  $\widehat{\langle X, X \rangle}_T^{(m_1)}$ , again delivers a marked improvement on the naive estimator,  $\widehat{\langle X, X \rangle}_T^{(b)}$ , and performs marginally better than the first-best estimator in [36],  $\widehat{\langle X, X \rangle}_T^{(s_1)}$ .

We also performed a Monte Carlo simulation for the Ornstein Uhlenbeck process given by:

$$dX_t = X_t dt + \sqrt{2\sigma_t^2} dB_t, \quad (39)$$

where also  $\sigma_t^2 = 0.01$ . Again we retain the same simulation setup and the results are displayed in Table III. The results are almost identical to that of the Brownian process, with the new estimator again outperforming other time-domain estimators.

	Sample bias	Sample variance	Sample RMSE
$\widehat{\langle X, X \rangle}_T^{(b)}$	$1.17 \times 10^{-2}$	$1.78 \times 10^{-8}$	$1.17 \times 10^{-2}$
$\widehat{\langle X, X \rangle}_T^{(s_1)}$	$6.69 \times 10^{-7}$	$2.66 \times 10^{-11}$	$5.20 \times 10^{-6}$
$\widehat{\langle X, X \rangle}_T^{(m_1)}$	$2.95 \times 10^{-7}$	$1.97 \times 10^{-11}$	$4.44 \times 10^{-6}$
$\widehat{\langle X, X \rangle}_T^{(m_2)}$	$5.09 \times 10^{-9}$	$6.76 \times 10^{-13}$	$8.22 \times 10^{-7}$
$\widehat{\langle X, X \rangle}_T^{(u)}$	$6.29 \times 10^{-9}$	$5.33 \times 10^{-13}$	$7.30 \times 10^{-7}$

TABLE III  
SIMULATION STUDY FOR THE ORNSTEIN UHLENBECK PROCESS.

	Sample bias	Sample variance	Sample RMSE
$\widehat{\langle X, X \rangle}_T^{(b)}$	$1.17 \times 10^{-3}$	$2.29 \times 10^{-9}$	$1.17 \times 10^{-3}$
$\widehat{\langle X, X \rangle}_T^{(s_1)}$	$1.00 \times 10^{-6}$	$4.51 \times 10^{-10}$	$2.13 \times 10^{-5}$
$\widehat{\langle X, X \rangle}_T^{(m_1)}$	$1.84 \times 10^{-7}$	$4.23 \times 10^{-10}$	$2.06 \times 10^{-5}$
$\widehat{\langle X, X \rangle}_T^{(m_2)}$	$4.80 \times 10^{-8}$	$2.42 \times 10^{-10}$	$1.55 \times 10^{-5}$
$\widehat{\langle X, X \rangle}_T^{(u)}$	$5.27 \times 10^{-8}$	$2.28 \times 10^{-10}$	$1.51 \times 10^{-5}$

TABLE IV  
SIMULATION STUDY FOR SHORTER SAMPLER LENGTH.

### C. Comparing estimators over shorter sample lengths

This section compares estimators for a shorter sample length which will reduce the benefit of subsampling due to the variance issues of small-length data but will also affect the variance of the multiscale ratio (*cf* Theorem 1).

The simulation setup is exactly the same as before (using the Heston model with the same parameters) except that  $T$ , the simulation length, is reduced by a factor of 10 to 0.1 days or 2340s. Before the results of the simulation are reported, it is of interest to see whether the frequency domain methods developed still model each process accurately. Figure 7 shows the calculated  $\hat{\sigma}_X^2$  and  $\hat{\sigma}_\varepsilon^2 |\sin(\pi \Delta t f_k)|^2$  (in white) together with the periodograms  $\hat{\mathcal{S}}_{kk}^{(X)}$  and  $\hat{\mathcal{S}}_{kk}^{(\varepsilon)}$  for one simulated path. The estimator still approximates the energy structure of the processes accurately. Figure 7 also shows the corresponding estimate of the multiscale ratio  $\hat{L}_k$  (in white) from this simulated path (together with  $\tilde{L}_k$ ) and the corresponding plot of  $\hat{L}_k \hat{\mathcal{S}}_{kk}^{(Y)}$ . The new estimator has removed the microstructure noise effect and has formed a good approximation of  $\hat{\mathcal{S}}_{kk}^{(X)}$ . The approximation of the periodograms is still accurate despite the shortening of available data.

Table IV displays the accuracy of the estimators over the 50,000 simulated paths. The first-best estimator in [36],  $\widehat{\langle X, X \rangle}_T^{(s_1)}$ , and the new estimator,  $\widehat{\langle X, X \rangle}_T^{(m_1)}$ , are once again comparable in performance and both estimates are close to the best attainable RMSE given by,  $\widehat{\langle X, X \rangle}_T^{(u)}$ , the realized integrated volatility on  $X_t$ .

### D. Comparing estimators with a low-noise process

This section compares estimators for smaller levels of microstructure noise. Reducing the microstructure noise will reduce the need to subsample. The first-best estimator in [36],  $\widehat{\langle X, X \rangle}_T^{(s_1)}$ , will have a higher sampling frequency and the new estimator will reduce its estimate of  $\hat{\sigma}_\varepsilon^2$  accordingly. For very small levels of noise, however, the first-best estimator will become zero, as the optimal number of samples becomes  $n$  (the highest available). This possibility is now examined, using the Heston model as before, with all parameters unchanged except the noise is reduced by a factor of 10, ie.  $\sigma_\varepsilon^2 = 0.00005^2$ . Note that the path length is kept at its original length of  $T = 1$  day.

Figure 8 shows the estimates of  $\hat{\sigma}_X^2$  and  $\hat{\sigma}_\varepsilon^2 |2 \sin(\pi \Delta t f_k)|^2$  (in white) along with the periodograms  $\hat{\mathcal{S}}_{kk}^{(Y)}$  and  $\hat{\mathcal{S}}_{kk}^{(\varepsilon)}$  for one simulated path along with the corresponding estimate of the multiscale ratio  $\hat{L}_k$  (in white) (plotted over the approximated  $\tilde{L}_k$ ) and the corresponding plot of  $\hat{L}_k \hat{\mathcal{S}}_{kk}^{(Y)}$ . The estimation method works well again; notice how the magnitude of the microstructure noise has been greatly reduced (the scale is now of order  $10^{-8}$  rather than  $10^{-6}$ ) causing the multiscale ratio  $L_k$  to be more tempered across the high frequencies than it was before, due

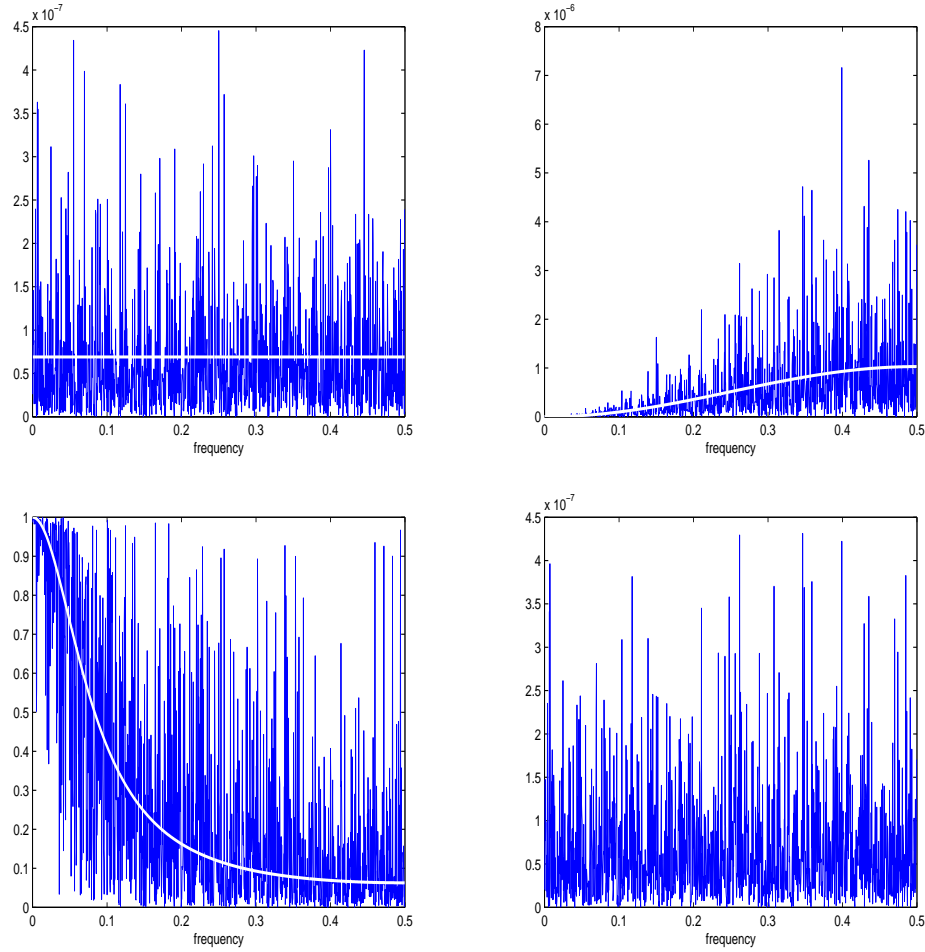


Fig. 7. A realisation of  $\widehat{S}_{kk}^{(X)}$  (top left), a realisation of  $\widehat{S}_{kk}^{(\epsilon)}$  (top right) with the Whittle estimates superimposed, the estimate of  $L_k$  (bottom left) with the Whittle estimate of  $L_k$  superimposed and the biased corrected estimator of  $S_{kk}^{(X)}$  using  $\widehat{L}_k \widehat{S}_{kk}^{(Y)}$  (bottom right). Notice the different scales in the four figures.

to the smaller microstructure noise. Nonetheless, the new estimator has still detected the smaller levels of noise in the data.

Table V reports on the results of 50,000 simulations performed as before. The first-best estimator of [36],  $\widehat{\langle X, X \rangle}_T^{(s_1)}$ , categorically failed for this model. This is due to the fact that the optimal number of samples was always equal to  $n$ , the total number of samples available. Therefore, the first-best estimator was always zero. The second-best estimator in [36], denoted by  $\widehat{\langle X, X \rangle}_T^{(s_2)}$ , was reasonably effective. This is simply an estimator that averages estimates calculated from sub-sampled paths at different starting points and is therefore asymptotically biased. The new estimator,  $\widehat{\langle X, X \rangle}_T^{(m_1)}$ , was remarkably robust, with RMSE very close to the RMSE of estimators based on the  $X_t$  process. The difference in performance between estimators using  $Y_t$  and estimators using  $X_t$  is expected to become smaller with less microstructure noise and this can be seen by the similar order RMSE errors between all estimators. Nevertheless, the new estimator was much closer in performance to the realized integrated volatility on  $X_t$  than it was to any other estimator on  $Y_t$ , a result that demonstrates the precision and robustness of this new estimator of integrated volatility.

### E. Correlated Noise

In this section we consider microstructure noise that is correlated. If this process is stationary, the noise process can be modelled as an MA process (as described in Section II-E), and the corresponding parameters can be estimated by maximising the multiscale Whittle likelihood using (17) and (32). Figure 9 shows the multiscale estimator



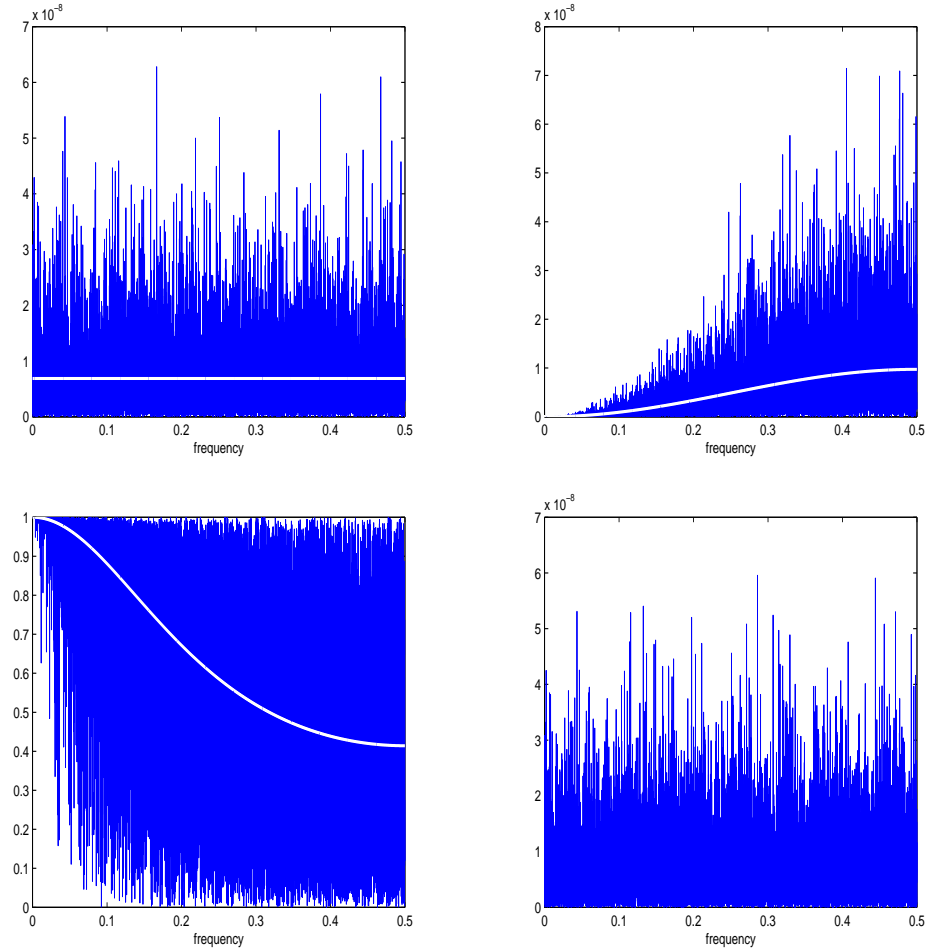


Fig. 8. A realisation of  $\widehat{S}_{kk}^{(X)}$  (top left), a realisation of  $\widehat{S}_{kk}^{(\epsilon)}$  (top right) with the Whittle estimates superimposed, the estimate of  $L_k$  (bottom left) with the Whittle estimate of  $L_k$  superimposed and the biased corrected estimator of  $\widehat{S}_{kk}^{(X)}$  using  $\widehat{L}_k \widehat{S}_{kk}^{(Y)}$  (bottom right). Notice the different scales in the four figures.

	Sample bias	Sample variance	Sample RMSE
$\widehat{\langle X, X \rangle}_T^{(b)}$	$1.17 \times 10^{-4}$	$2.11 \times 10^{-10}$	$1.18 \times 10^{-4}$
$\widehat{\langle X, X \rangle}_{T(s_2)}$	$3.53 \times 10^{-6}$	$1.00 \times 10^{-9}$	$3.19 \times 10^{-5}$
$\widehat{\langle X, X \rangle}_{T(m_1)}$	$7.63 \times 10^{-9}$	$2.12 \times 10^{-10}$	$1.46 \times 10^{-5}$
$\widehat{\langle X, X \rangle}_{T(m_2)}$	$7.91 \times 10^{-9}$	$2.06 \times 10^{-10}$	$1.44 \times 10^{-5}$
$\widehat{\langle X, X \rangle}_T^{(u)}$	$9.83 \times 10^{-9}$	$2.05 \times 10^{-10}$	$1.43 \times 10^{-5}$

TABLE V  
SIMULATION STUDY FOR LOWER MARKET MICROSTRUCTURE NOISE.

applied to the Heston Model (with the same parameters as before) with a microstructure noise that follows an MA(6) process (parameters given in the caption). The Whittle estimates (in white) form a good approximation of  $\widehat{S}_{kk}^{(X)}$  and  $\widehat{S}_{kk}^{(\epsilon)}$  despite the more complicated nuisance structure. The corresponding estimate of the multiscale ratio  $\widehat{L}_k$  (in white) therefore removes energy from the correct frequencies and the corresponding plot of  $\widehat{L}_k \widehat{S}_{kk}^{(Y)}$  is a good approximation of  $\widehat{S}_{kk}^{(X)}$ . This is the same noise process and Itô process for which we calculated the optimal smoothing window in section II-E, and the trough in the noise at about  $f = 0.42$  corresponds to the oscillations in the kernel plotted in Figure 1.

If the length of the MA( $p$ ) process is unknown, then  $p$  can be determined using (33). In Table VI we show an

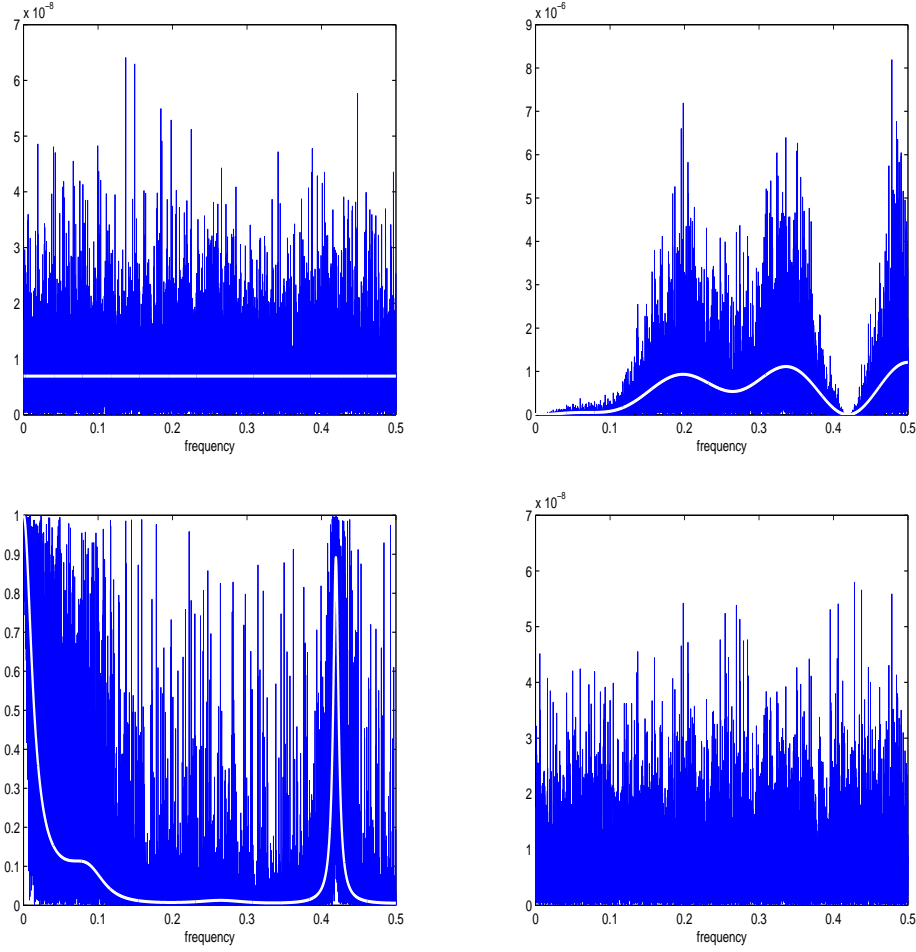


Fig. 9. A realisation of  $\widehat{S}_{kk}^{(X)}$  (top left), a realisation of  $\widehat{S}_{kk}^{(\epsilon)}$  (top right) with the Whittle estimates superimposed, the estimate of  $L_k$  (bottom left) with the Whittle estimate of  $L_k$  superimposed and the biased corrected estimator of  $\widehat{S}_{kk}^{(X)}$  using  $\widehat{L}_k \widehat{S}_{kk}^{(Y)}$  (bottom right). In this example we use an MA(6) with  $\theta_1 = 0.5$ ,  $\theta_2 = -0.1$ ,  $\theta_3 = -0.1$ ,  $\theta_4 = 0.2$ ,  $\theta_5 = 0$  and  $\theta_6 = 0.4$ . Notice the different scales in the four figures.

MA( $p$ )	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_5$	$\theta_6$	$\theta_7$	$\theta_8$	AICC
$p = 1$	0.935								$-3.208490 \times 10^5$
$p = 2$	0.624	-0.445							$-3.239947 \times 10^5$
$p = 3$	0.658	-0.459	-0.046						$-3.240000 \times 10^5$
$p = 4$	0.806	-0.603	-0.101	0.410					$-3.262427 \times 10^5$
$p = 5$	0.813	-0.606	-0.101	0.411	-0.008				$-3.262416 \times 10^5$
$p = 6$	0.815	-0.604	-0.097	0.420	-0.003	0.000			$-3.262409 \times 10^5$
$p = 7$	0.807	-0.613	-0.114	0.413	0.002	-0.002	-0.005		$-3.262402 \times 10^5$
$p = 8$	0.817	-0.614	-0.128	0.427	0.005	0.011	-0.009	-0.017	$-3.262384 \times 10^5$

TABLE VI

VALUES OF  $\theta$  FOUND BY MODELLING THE NOISE PROCESS AS AN MA( $p$ ) PROCESS FOR  $p = 1, \dots, 8$ . MODEL CHOICE METHODS (AICC) ARE USED TO SELECT WHICH PROCESS TO MODEL THE NOISE BY, IN THIS CASE THE AICC IS MINIMISED BY SELECTING AN MA(4) WITH THE GIVEN PARAMETERS. THE TRUE NOISE IS INDEED AN MA(4) PROCESS (WITH PARAMETERS  $\theta_1 = 0.8$ ,  $\theta_2 = -0.6$ ,  $\theta_3 = 0.1$ ,  $\theta_4 = 0.4$ ).

example with  $p = 4$  with parameters  $\theta_1 = 0.8$ ,  $\theta_2 = -0.6$ ,  $\theta_3 = 0.1$ ,  $\theta_4 = 0.4$ . Clearly  $p = 4$  is identified as the best fitting model yielding near to perfect estimates of the noise parameters. The estimator is therefore robust to removing the effect of microstructure noise when this process is correlated (and stationary), even if the length of the MA( $p$ ) process is not explicitly known.

We also tested our estimator using Monte Carlo simulations in [31] for a variety of MA(1) processes and the

results showed a significant reduction in error compared with not only the naive estimator, but also the estimators based on a white-noise assumption. Furthermore, the adjusted multiscale estimator performed almost identically to our multiscale estimator when we set  $\theta_1 = 0$  and recovered a white-noise process, meaning the loss in precision from searching for a parameter unnecessarily was negligible (as to be expected for  $q \ll N$ ). Notice also that in Table VI there appears to be little loss in precision from estimating more parameters in the MA(4) process than is required as  $\theta_p$  for  $p > 4$  is always estimated to be very close to zero. This further demonstrates the robustness and precision of our estimation technique.

#### IV. CONCLUSIONS

The problem of estimating the integrated stochastic volatility of an Itô process from noisy observations was studied in this paper. Unlike most previous works on this problem, see [26], [36], the method for estimating the integrated volatility developed in this paper is based on the frequency domain representation of both the Itô process and the noisy observations. The integrated volatility can be represented as a summation of variation in the process of interest over all frequencies (or scales). In our estimator we adjust the raw sample variance at each frequency. Such an estimator is truly multiscale, as it corrects the estimated energy directly at every scale. In other words, the estimator is debiased *locally at each frequency*, rather than globally.

To estimate the degree of scale separation in the data we used the Whittle likelihood, and quantified the noise contribution by the multiscale ratio. Various properties of the multiscale estimator were determined, see Theorems 1 and 2. As was illustrated by the set of examples, our estimator performs extremely well on data simulated from the Heston model, and is competitive with the methods proposed by [36], under varying signal-to-noise and sampling scenarios. The proposed estimator is truly multiscale in nature and adapts automatically to the degree of noise contamination of the data, a clear strength. It is also easily implemented and computationally efficient.

The new estimator for the integrated stochastic volatility can be written as

$$\widehat{\langle X, X \rangle} = \sum_u \ell_{-u} \sum_k (X_{t_{k-u}} - X_{t_{k-u-1}}) (X_{t_k} - X_{t_{k-1}}),$$

where the kernel  $\ell_u$  is given by (28). We can compare this estimator with kernel estimators, see [10]. There the estimated increment square  $\Delta X_t^2$  is locally smoothed to estimate the diffusion coefficient using a kernel function,  $K(\cdot)$ . Contrary to this approach we estimate the integrated volatility by smoothing the estimated autocovariance of  $\Delta X_{t_j}$ . In particular, we use a data-dependent choice of smoothing window. We show that, from a minimum bias perspective, using a Laplace window to smooth is optimal. This data-dependent choice of smoothing window becomes more interesting after relaxing the assumptions on the noise process, and treating correlated observation error.

Inference procedures implemented in the frequency domain are still very underdeveloped for problems with a multiscale structure. The modern data deluge has caused an excess of high frequency observations in a number of application areas, for example finance and molecular dynamics. More flexible models could also be used for the high frequency nuisance structure. In this paper we have introduced a new frequency domain based estimator and applied it to a relatively simple problem, namely the estimation of the integrated stochastic volatility, for data contaminated by high frequency noise. There are many extensions and potential applications of the new estimator. Here we list a few which seem interesting to us and which are currently under investigation.

- Study parameter estimation for noisily observed SDEs which are driven by more general noise processes, for example Lévy processes.
- Application of the new estimator to the problem of statistical inference for fast/slow systems of SDEs, of the type studied in [24], [26].
- Study the combined effects of high-frequency and multiscale structure in the data. A first step in this direction was taken in [7].

#### REFERENCES

- [1] Y. AIT-SAHALIA, P. A. MYKLAND, AND L. ZHANG, *How often to sample a continuous-time process in the presence of market microstructure noise*, Rev. Financ. Studies, 18 (2005), pp. 351–416.
- [2] E. BARUCCI AND R. RENO, *On measuring volatility of diffusion processes with high frequency data*, Economics Letters, 74 (2002), pp. 371–378.

- [3] J. BERAN, *Statistics for long-memory Processes*, Chapman and Hall, London, 1994.
- [4] PETER BLOOMFIELD, *Fourier Analysis of Time Series*, Wiley-IEEE, New York, 2004.
- [5] D. BRILLINGER, *Time series: data analysis and theory*, Society for Industrial and Applied Mathematics, Philadelphia, USA, 2001.
- [6] P. J. BROCKWELL AND R. A. DAVIS, *Time Series: Theory and Methods*, Springer-Verlag, New York, 1991.
- [7] C. J. COTTER AND G. A. PAVLIOTIS, *Estimating Eddy diffusivities from Lagrangian Observations*, Preprint, 2009.
- [8] K. O DZHAMPARIDZE AND A. M. YAGLOM, *Spectrum parameter estimation in time series analysis*, in Developments in Statistics, PR Krishnaiah, ed., vol. 4, New York: Academic Press., 1983, pp. 1–181.
- [9] J. FAN AND Y. WANG, *Multi-scale jump and volatility analysis for high-frequency financial data*, J. of the American Statistical Association, 102 (2007), pp. 1349–1362.
- [10] D. FLORENS-ZMIROU, *On estimating the diffusion coefficient from discrete observations*, Journal of Applied Probability, 30 (1993), pp. 790–804.
- [11] J.-P. FOUQUE, G. PAPANICOLAOU, K. R. SIRCAR, AND K. SOLNA, *Short time-scale in SP-500 volatility*, J. Comp. Finance, 6 (2003), pp. 1–23.
- [12] D. GIVON, R. KUPFERMAN, AND A. M. STUART, *Extracting macroscopic dynamics: model problems and algorithms*, Nonlinearity, 17 (2004), pp. R55–R127.
- [13] A. GRIFFA, K. OWENSA, L. PITERBARG, AND B. ROZOVSKII, *Estimates on turbulence parameters from lagrangian data using a stochastic particle model*, J. Marine Research, 53 (1995), pp. 371–401.
- [14] P. R. HANSEN AND A. LUNDE, *A forecast comparison of volatility models: does anything beat a GARCH(1,1)?*, Journal of Applied Econometrics, 20 (2005), pp. 873–889.
- [15] ———, *Realized variance and market microstructure noise*, J. Business & Economic Statistics, 24 (2006), pp. 127–161.
- [16] S. L. HESTON, *A closed form solution for options with stochastic volatility with applications to bond and currency options*, Review of Financial Studies, 6 (1993), pp. 327–343.
- [17] I. HORENKO, C. HARTMANN, C. SCHÜTTE AND F. NOE, *Data-based parameter estimation of generalized multidimensional Langevin processes*, Physical Review E, 76 (2007), no. 016706.
- [18] I. HORENKO AND C. SCHÜTTE, *Likelihood-based estimation of multidimensional Langevin Models and its Application to biomolecular dynamics*, Multiscale Modeling and Simulation, 7 (2008), 731–773.
- [19] I. KARATZAS AND S. E. SHREVE, *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York, 1991.
- [20] Y. A. KUTOYANTS, *Statistical inference for ergodic diffusion processes*, Springer Series in Statistics, London, 2004.
- [21] P. MALLIAVIN AND M. E. MANCINO, *Fourier series for measurements of multivariate volatilities*, Finance and Stochastics, 6 (2002), pp. 49–61.
- [22] M. E. MANCINO AND S. SANFELICI, *Robustness of Fourier estimators of integrated volatility in the presence of microstructure noise*, tech. report, University of Firenze, 2006.
- [23] F. NEESER AND J. MASSEY, *Proper complex random processes with applications to information theory*, IEEE Trans. Info. Theory, 39 (1993), 1293–1302.
- [24] T. PAPAVALIOU, G. A. PAVLIOTIS, AND A.M. STUART, *Maximum likelihood estimation for multiscale diffusions*, 2008.
- [25] G. A. PAVLIOTIS, Y. POKERN, AND A.M. STUART, *Parameter estimation for multiscale diffusions: an overview*, 2008.
- [26] G. A. PAVLIOTIS AND A. M. STUART, *Parameter estimation for multiscale diffusions*, J. Stat. Phys., 127 (2007), pp. 741–781.
- [27] D. B. PERCIVAL AND A. T. WALDEN, *Spectral Analysis for Physical Applications*, Cambridge University Press, Cambridge, UK, 1993.
- [28] B. PICINBONO, *Second-Order Complex Random Vectors and Normal Distributions*, IEEE Trans. Signal Proc., 44 (1996), pp. 2637–2640.
- [29] R. RENÒ, *Volatility estimate via Fourier analysis*, PhD thesis, Scuola Normale Superiore, 2005.
- [30] ———, *Nonparametric estimation of the diffusion coefficient of stochastic volatility models*, Econometric Theory, 24 (2008), pp. 1174–1206.
- [31] A. SYKULSKI, S. OLHEDE, AND G. A. PAVLIOTIS, *High frequency variability and microstructure bias*, in Inference and Estimation in Probabilistic Time-Series Models, D. Barber, A. T. Cemgil, and S. Chiappa, eds., Cambridge, UK, 2008, Issac Newton Institute for Mathematical Sciences. Proceedings available at <http://www.newton.ac.uk/programmes/SCH/schw05-papers.pdf>.
- [32] R. S. TSAY, *Analysis of Financial Time Series*, John Wiley & Sons, New York, USA, 2005.
- [33] L. WASSERMAN, *All of Nonparametric Statistics*, Springer, Berlin, 2007.
- [34] P. WHITTLE, *Estimation and information in stationary time series*, Arkiv för Matematik, 2 (1953), pp. 423–434.
- [35] P. WHITTLE, *Gaussian estimation in stationary time series*, Bull. Inst. Statist. Inst., 39 (1962), pp. 105–129.
- [36] L. ZHANG, P. A. MYKLAND, AND Y. AIT-SAHALIA, *A tale of two time scales: Determining integrated volatility with noisy high-frequency data*, J. Am. Stat. Assoc., 100 (2005), pp. 1394–1411.

#### A. PROOF OF THEOREM 1

Let the true value of  $\sigma$  be denoted  $\sigma^*$ . We differentiate the multiscale energy likelihood function (18) with respect to  $\sigma$  to obtain

$$\begin{aligned}\dot{\ell}_X(\sigma) &= \frac{\partial \ell(\sigma)}{\partial \bar{\sigma}_X^2} = - \sum_{k=1}^{N/2-1} \frac{1}{\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2} + \sum_{k=1}^{N/2-1} \frac{\hat{\mathcal{S}}_{kk}^{(Y)}}{(\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2)^2} \\ \dot{\ell}_\varepsilon(\sigma) &= \frac{\partial \ell(\sigma)}{\partial \sigma_\varepsilon^2} = - \sum_{k=1}^{N/2-1} \frac{|2 \sin(\pi f_k \Delta t)|^2}{\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2} + \sum_{k=1}^{N/2-1} \frac{|2 \sin(\pi f_k \Delta t)|^2 \hat{\mathcal{S}}_{kk}^{(Y)}}{(\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2)^2}.\end{aligned}$$

To remove implicit  $\Delta t$  dependence we let  $\tau_X = \bar{\sigma}_X^2 / \Delta t$ , and denote derivatives with respect to  $\tau_X$  by subscript  $\tau$ . Then  $\dot{\ell}_\tau(\hat{\sigma}) = \Delta t \dot{\ell}_X(\hat{\sigma})$ , and so on. We calculate the expectation and variance of the score functions evaluated at  $\sigma^*$ , and find that the bias of  $\hat{\tau}_X$  is order  $\mathcal{O}(\Delta t^{1/2} \log(\Delta t))$  and the bias of  $\hat{\sigma}_\varepsilon^2$  is order  $\mathcal{O}(\Delta t^2 \log(\Delta t))$ . These contributions become negligible, and are of lesser importance compared to the variance.

To show large sample properties we Taylor expand the multiscale likelihood with  $\hat{\sigma}$  corresponding to the estimated maximum likelihood, and  $\sigma'$  is lying between  $\hat{\sigma}$  and  $\sigma^*$ . Then

$$\begin{aligned}\dot{\ell}_\tau(\hat{\sigma}) &= \dot{\ell}_\tau(\sigma^*) + \ddot{\ell}_{\tau\tau}(\sigma') [\hat{\sigma}_X^2 - \sigma_X^{*2}] / \Delta t + \ddot{\ell}_{\tau\varepsilon}(\sigma') [\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^{*2}] \\ \dot{\ell}_\varepsilon(\hat{\sigma}) &= \dot{\ell}_\varepsilon(\sigma^*) + \ddot{\ell}_{\varepsilon\tau}(\sigma') [\hat{\sigma}_X^2 - \sigma_X^{*2}] / \Delta t + \ddot{\ell}_{\varepsilon\varepsilon}(\sigma') [\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^{*2}].\end{aligned}$$

We note with the observed Fisher information

$$\mathbf{F} = \begin{bmatrix} \ddot{\ell}_{\tau\tau}(\sigma') & \ddot{\ell}_{\tau\varepsilon}(\sigma') \\ \ddot{\ell}_{\varepsilon\tau}(\sigma') & \ddot{\ell}_{\varepsilon\varepsilon}(\sigma') \end{bmatrix}$$

that

$$\begin{pmatrix} (\hat{\sigma}_X^2 - \sigma_X^{*2}) / \Delta t \\ \hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^{*2} \end{pmatrix} = \mathbf{F}^{-1} \begin{pmatrix} \dot{\ell}_\tau(\hat{\sigma}) - \dot{\ell}_\tau(\sigma^*) \\ \dot{\ell}_\varepsilon(\hat{\sigma}) - \dot{\ell}_\varepsilon(\sigma^*) \end{pmatrix}. \quad (\text{A-40})$$

We henceforth ignore the term  $J_k^{(\mu)} = J_k^{(X)} - \tilde{J}_k^{(X)}$  as this will not contribute to leading order, and write  $J_k^{(X)}$  where formally we would write  $\tilde{J}_k^{(X)}$  or  $J_k^{(X)}$ . We can observe the suitability of this directly from (18) and use bounds for  $J_k^{(\mu)}$ , where we could formally apply these to get bounds on each derivative of  $l(\sigma)$  (note that we cannot differentiate bounds). To avoid needless technicalities, the details of this approach will not be reported. To leading order

$$\begin{aligned}\text{var}(\dot{\ell}_\tau(\sigma)) &= \sum_{l=1}^{N/2-1} \sum_{k=1}^{N/2-1} \frac{\Delta t^2 \text{cov}(\hat{\mathcal{S}}_{kk}^{(Y)}, \hat{\mathcal{S}}_{ll}^{(Y)})}{\left(\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2\right)^2 \left(\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_l \Delta t)|^2\right)^2} \\ \text{var}(\dot{\ell}_\varepsilon(\sigma)) &= \sum_{l=1}^{N/2-1} \sum_{k=1}^{N/2-1} \frac{|2 \sin(\pi f_k \Delta t)|^2 |2 \sin(\pi f_l \Delta t)|^2 \text{cov}(\hat{\mathcal{S}}_{kk}^{(Y)}, \hat{\mathcal{S}}_{ll}^{(Y)})}{\left(\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2\right)^2 \left(\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_l \Delta t)|^2\right)^2} \\ \text{cov}(\dot{\ell}_\tau(\sigma), \dot{\ell}_\varepsilon(\sigma)) &= \sum_{l=1}^{N/2-1} \sum_{k=1}^{N/2-1} \frac{\Delta t |2 \sin(\pi f_l \Delta t)|^2 \text{cov}(\hat{\mathcal{S}}_{kk}^{(Y)}, \hat{\mathcal{S}}_{ll}^{(Y)})}{\left(\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2\right)^2 \left(\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_l \Delta t)|^2\right)^2}.\end{aligned}$$

We now need to calculate  $\text{cov}(\hat{\mathcal{S}}_{kk}^{(Y)}, \hat{\mathcal{S}}_{ll}^{(Y)})$  which is

$$\begin{aligned}\text{cov}(\hat{\mathcal{S}}_{kk}^{(Y)}, \hat{\mathcal{S}}_{ll}^{(Y)}) &= \mathbb{E} \left\{ J_k^{(Y)} [J_k^{(Y)}]^* [J_l^{(Y)}]^* J_l^{(Y)} \right\} - \mathbb{E} \left\{ \hat{\mathcal{S}}_{kk}^{(Y)} \right\} \mathbb{E} \left\{ \hat{\mathcal{S}}_{ll}^{(Y)} \right\} \\ &= \rho_{kl}^{(Y)} \mathcal{S}_{kk}^{(Y)} \mathcal{S}_{ll}^{(Y)}.\end{aligned} \quad (\text{A-41})$$

Furthermore

$$\begin{aligned}& \mathbb{E} \left\{ J_k^{(Y)} [J_k^{(Y)}]^* [J_l^{(Y)}]^* J_l^{(Y)} \right\} - \mathbb{E} \left\{ J_k^{(Y)} [J_k^{(Y)}]^* \right\} \mathbb{E} \left\{ [J_l^{(Y)}]^* J_l^{(Y)} \right\} \\ &= \mathbb{E} \left\{ (J_k^{(X)} + J_k^{(\varepsilon)}) [(J_k^{(X)} + J_k^{(\varepsilon)})]^* [(J_l^{(X)} + J_l^{(\varepsilon)})]^* (J_l^{(X)} + J_l^{(\varepsilon)}) \right\} \\ &\quad - \mathbb{E} \left\{ J_k^{(Y)} [J_k^{(Y)}]^* \right\} \mathbb{E} \left\{ [J_l^{(Y)}]^* J_l^{(Y)} \right\} \\ &= \text{cov} \left\{ \hat{\mathcal{S}}_{kk}^{(X)}, \hat{\mathcal{S}}_{ll}^{(X)} \right\} + \text{cov} \left\{ \hat{\mathcal{S}}_{kk}^{(\varepsilon)}, \hat{\mathcal{S}}_{ll}^{(\varepsilon)} \right\} + \mathcal{S}_{kl}^{(X)} \mathcal{S}_{kl}^{(\varepsilon)*} + \mathcal{S}_{kl}^{(X)*} \mathcal{S}_{kl}^{(\varepsilon)}.\end{aligned}$$

We therefore need to calculate the individual terms of this expression. We note

$$\text{cov} \left\{ \hat{\mathcal{S}}_{kk}^{(\varepsilon)}, \hat{\mathcal{S}}_{ll}^{(\varepsilon)} \right\} = \delta_{kl} [\mathcal{S}_{kk}^{(\varepsilon)}]^2, \quad \mathcal{S}_{kl}^{(X)} \mathcal{S}_{kl}^{(\varepsilon)*} + \mathcal{S}_{kl}^{(X)*} \mathcal{S}_{kl}^{(\varepsilon)} = 2\delta_{kl} \mathcal{S}_{kk}^{(X)} \mathcal{S}_{kk}^{(\varepsilon)}.$$

Then it follows

$$\text{cov} \left\{ \hat{\mathcal{S}}_{kk}^{(Y)}, \hat{\mathcal{S}}_{ll}^{(Y)} \right\} = \text{cov} \left\{ \hat{\mathcal{S}}_{kk}^{(X)}, \hat{\mathcal{S}}_{ll}^{(X)} \right\} + \delta_{kl} [\mathcal{S}_{kk}^{(\varepsilon)}]^2 + 2\delta_{kl} \mathcal{S}_{kk}^{(X)} \mathcal{S}_{kk}^{(\varepsilon)}. \quad (\text{A-42})$$

We therefore only need to worry about  $\text{cov} \left\{ \widehat{\mathcal{S}}_{kk}^{(X)}, \widehat{\mathcal{S}}_{ll}^{(X)} \right\}$ . We need

$$\begin{aligned} \mathbb{E} \left\{ J_k^{(X)} [J_k^{(X)}]^* [J_l^{(X)}]^* J_l^{(X)} \right\} &= \frac{1}{N^2} \mathbb{E} \left\{ \sum_{n=1}^N \int_{(n-1)\Delta t}^{n\Delta t} \sigma_s dW_s e^{-2i\pi \frac{kn}{N}} \right. \\ &\quad \times \sum_{p=1}^N \int_{(p-1)\Delta t}^{p\Delta t} \sigma_t dW_t e^{2i\pi \frac{kp}{N}} \sum_{m=1}^N \int_{(m-1)\Delta t}^{m\Delta t} \sigma_u dW_u e^{-2i\pi \frac{lm}{N}} \sum_{w=1}^N \int_{(w-1)\Delta t}^{w\Delta t} \sigma_v dW_v e^{2i\pi \frac{lw}{N}} \left. \right\} \\ &=: \frac{1}{N^2} \sum_{n=1}^N \sum_{p=1}^N \sum_{m=1}^N \sum_{\rho=1}^N \left( e_{kn} e_{kp}^* e_{\ell m}^* e_{\ell \rho} \mathbb{E} \left( M_n M_p M_m M_\rho \right) \right), \end{aligned}$$

where  $M_n := \int_{(n-1)\Delta t}^{n\Delta t} \sigma_s dW_s$  and  $e_{kn} := e^{-2i\pi \frac{kn}{N}}$ . Since Brownian motion has independent increments, we have that  $\mathbb{E} \left( M_n M_p M_m M_\rho \right) = EM_n^4$  if  $n = p = m = \rho$ ,  $\mathbb{E} \left( M_n M_k M_m M_\rho \right) = EM_n^2 EM_k^2$  if  $n = k$ ,  $m = \rho$  and  $\mathbb{E} \left( M_n M_p M_m M_\rho \right) = 0$ , otherwise. Consequently,

$$\begin{aligned} \mathbb{E} \left\{ J_k^{(X)} [J_k^{(X)}]^* [J_l^{(X)}]^* J_l^{(X)} \right\} &= \frac{1}{N^2} \sum_{n=1}^N EM_n^4 + \frac{1}{N^2} \left( \sum_{n=1}^N EM_n^2 \right)^2 \\ &\quad + \frac{1}{N^2} \sum_{n=1}^N \sum_{p=1}^N e_{kn} e_{\ell n}^* e_{kp}^* e_{\ell p} EM_n^2 EM_p^2 \\ &\quad + \frac{1}{N^2} \sum_{n=1}^N \sum_{p=1}^N e_{kn} e_{\ell p}^* e_{kp}^* e_{\ell n} EM_n^2 EM_p^2 \end{aligned}$$

We use standard bounds on moments of stochastic integrals [19] to obtain the bound

$$\frac{1}{N^2} \sum_{n=1}^N EM_n^4 \leq \frac{1}{N^2} \sum_{n=1}^N 36\Delta t \int_{(n-1)\Delta t}^{n\Delta t} \mathbb{E} \sigma_s^4 ds \leq C(\Delta t)^3,$$

since, by assumption,  $\mathbb{E} \sigma_s^4 = \mathcal{O}(1)^2$ . We have:

$$\begin{aligned} \rho_{kl}^{(X)} \mathcal{S}_{kk}^{(X)} \mathcal{S}_{ll}^{(X)} &= \mathbb{E} \left\{ J_k^{(X)} [J_k^{(X)}]^* [J_l^{(X)}]^* J_l^{(X)} \right\} - \mathbb{E} \left| J_k^{(X)} \right|^2 \mathbb{E} \left| J_l^{(X)} \right|^2 \\ &= \frac{1}{N^2} \int_0^T \int_0^T \left( \cos(2\pi(k+l)(\frac{s-u}{T})) + \cos(2\pi(k-l)(\frac{s-u}{T})) \right) \\ &\quad \times \mathbb{E} \{ \sigma_s^2 \} \mathbb{E} \{ \sigma_u^2 \} ds du + \mathcal{O}((\Delta t)^3) \\ &= \frac{1}{2N^2} \int_0^T \int_0^T \mathbb{E} \{ \sigma_s^2 \} \mathbb{E} \{ \sigma_u^2 \} \left( e^{2i\pi(k+l)(\frac{s-u}{T})} + e^{-2i\pi(k+l)(\frac{s-u}{T})} \right. \\ &\quad \left. + e^{2i\pi(k-l)(\frac{s-u}{T})} + e^{-2i\pi(k-l)(\frac{s-u}{T})} \right) ds du + \mathcal{O}((\Delta t)^3) \\ &= \frac{1}{2N^2} \left( \Sigma(-\frac{k+l}{T}) \Sigma(\frac{k+l}{T}) + \Sigma(\frac{k+l}{T}) \Sigma(-\frac{k+l}{T}) \right. \\ &\quad \left. + \Sigma(-\frac{k-l}{T}) \Sigma(\frac{k-l}{T}) + \Sigma(\frac{k-l}{T}) \Sigma(-\frac{k-l}{T}) \right) + \mathcal{O}((\Delta t)^3). \end{aligned}$$

Since  $\mathbb{E} \sigma_t^2$  is a smooth function of time we can bound the decay of  $\Sigma(f) \propto \frac{1}{f}$  so that:

$$\rho_{kl}^{(X)} \mathcal{S}_{kk}^{(X)} \mathcal{S}_{ll}^{(X)} = \Delta t^2 \left( \mathcal{O} \left( \frac{1}{(k+l)^2} \right) + \mathcal{O} \left( \frac{1}{(k-l)^2} \right) \right). \quad (\text{A-43})$$

<sup>2</sup> $C$  in this paper denotes a generic constant, rather than the same constant.

We combine the foregoing calculations with (A-42)

$$\begin{aligned} \text{var} \left\{ \widehat{\mathcal{S}}_{kk}^{(Y)} \right\} &= \left( \mathcal{S}_{kk}^{(X)} + \mathcal{S}_{kk}^{(\varepsilon)} \right)^2. \\ \text{var} \left( \dot{\ell}_\tau(\widehat{\sigma}) \right) &= \sum_{l=1}^{N/2-1} \sum_{k=1}^{N/2-1} \frac{\Delta t^2 \text{cov} \left( \widehat{\mathcal{S}}_{kk}^{(Y)}, \widehat{\mathcal{S}}_{ll}^{(Y)} \right)}{\left( \bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2 \right)^2 \left( \bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_l \Delta t)|^2 \right)^2}. \end{aligned} \quad (\text{A-44})$$

We note that

$$\text{cov} \left( \widehat{\mathcal{S}}_{kk}^{(Y)}, \widehat{\mathcal{S}}_{ll}^{(Y)} \right) = \rho_{kl}^{(X)} \mathcal{S}_{kk}^{(X)} \mathcal{S}_{ll}^{(X)} + \delta_{kl} \left[ \mathcal{S}_{ll}^{(\varepsilon)} \right]^2 + 2\delta_{kl} \mathcal{S}_{kk}^{(X)} \mathcal{S}_{kk}^{(\varepsilon)}.$$

Thus it follows that:

$$\begin{aligned} \text{var} \left( \dot{\ell}_\tau(\widehat{\sigma}) \right) &= \sum_{l=1}^{N/2-1} \frac{\Delta t^2}{\left( \bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_l \Delta t)|^2 \right)^2} + C + \mathcal{O}(\log(\Delta t) \Delta t^{-1/4}) \\ &= \mathcal{O}(\Delta t^{-1/2}) + C + \mathcal{O}(\log(\Delta t) \Delta t^{-1/4}). \end{aligned} \quad (\text{A-45})$$

The extra order terms acknowledge potential effects from the drift. We need to establish the size of  $C$ . Using (A-42) we find that:

$$\begin{aligned} |C| &\leq \sum_{l \neq k}^{N/2-1} \frac{\Delta t^4 C_2 ((k+l)^{-2} + (k-l)^{-2})}{\left( \bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2 \right)^2 \left( \bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_l \Delta t)|^2 \right)^2} \\ &\leq 2 \sum_{k=1}^{N/2-1} \sum_{\tau=1}^k \frac{\Delta t^4 C_2 ((2k-\tau)^{-2} + \tau^{-2})}{\left( \bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_{k-\tau} \Delta t)|^2 \right)^2 \left( \bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2 \right)^2} \\ &\sim 2 \sum_{k=1}^{N/2-1} \sum_{\tau=1}^k \frac{C_2 ((2k-\tau)^{-2} + \tau^{-2})}{\left( \tau_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_{k-\tau} \Delta t)|^2 / \Delta t \right)^2 \left( \tau_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2 / \Delta t \right)^2} \\ &= \mathcal{O}(\log(\Delta t)). \end{aligned}$$

This is negligible in size compared to  $\Delta t^{-1/2}$ . Similar calculations can bound contributions from the off diagonals in the other two calculations. Also as  $\bar{\sigma}_X^2 = \tau_X \Delta t$

$$\begin{aligned} -\text{E} \left\{ \ddot{\ell}_{\tau\tau}(\sigma) \right\} &= \sum_{k=1}^{N/2-1} \frac{\Delta t^2}{\left( \bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2 \right)^2} + \mathcal{O}(\log(\Delta t)) = \mathcal{O}(\Delta t^{-1/2}) \\ -\text{E} \left\{ \ddot{\ell}_{\varepsilon\varepsilon}(\sigma) \right\} &= \sum_{k=1}^{N/2-1} \frac{|2 \sin(\pi f_k)|^4}{\left( \bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2 \right)^2} + \mathcal{O}(\log(\Delta t)) = \mathcal{O}(\Delta t^{-1}) \\ -\text{E} \left\{ \ddot{\ell}_{\tau\varepsilon}(\sigma) \right\} &= \sum_{k=1}^{N/2-1} \frac{\Delta t |2 \sin(\pi f_k)|^2}{\left( \bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2 \right)^2} + \mathcal{O}(\log(\Delta t)) = \mathcal{O}(\Delta t^{-1/2}). \end{aligned} \quad (\text{A-46})$$

The order terms follow from usual spectral theory on the white noise process, as well as bounds on  $J_k^{(\mu)}$ . We can also by considering the variance of the observed Fisher information deduce that renormalized versions of the entries of the observed Fisher information converge in probability to a constant, or

$$\text{diag}(\Delta t^{1/4}, \Delta t^{1/2}) \mathbf{F} \text{diag}(\Delta t^{1/4}, \Delta t^{1/2}) \longrightarrow \mathcal{F},$$

and thus using Slutsky's theorem we can deduce that:

$$\text{diag}(\Delta t^{-1/4}, \Delta t^{-1/2}) \left[ \begin{pmatrix} \bar{\sigma}_X^2 / \Delta t \\ \bar{\sigma}_\varepsilon^2 \end{pmatrix} - \begin{pmatrix} \bar{\sigma}_X^{*2} / \Delta t \\ \bar{\sigma}_\varepsilon^{*2} \end{pmatrix} \right] \text{diag}(\Delta t^{-1/4}, \Delta t^{-1/2}) \xrightarrow{L} N(\mathbf{0}, \mathcal{F}^{-1}),$$

where the entries of  $\mathcal{F}$  can be found from (A-46), (A-44) and (A-45), and

$$\begin{aligned} & \text{diag}(\Delta t^{-1/4}, \Delta t^{-1/2}) \text{var} \left\{ \left[ \left( \frac{\hat{\sigma}_X^2}{\hat{\sigma}_\varepsilon^2} / \Delta t \right) - \left( \frac{\bar{\sigma}_X^{*2}}{\sigma_\varepsilon^{*2}} / \Delta t \right) \right] \right\} \text{diag}(\Delta t^{-1/4}, \Delta t^{-1/2}) \\ &= \text{diag}(\Delta t^{-1/4}, \Delta t^{-1/2}) \mathbf{F}^{-1} \mathbf{F} \mathbf{F}^{-1} \text{diag}(\Delta t^{-1/4}, \Delta t^{-1/2}) \\ &= \text{diag}(\Delta t^{-1/4}, \Delta t^{-1/2}) \mathbf{F}^{-1} \text{diag}(\Delta t^{-1/4}, \Delta t^{-1/2}) \longrightarrow \mathcal{F}^{-1}. \end{aligned}$$

We have

$$\mathcal{F} = \begin{pmatrix} \frac{T}{\sigma_\varepsilon 16} \frac{1}{\tau_X^{3/2}} & 0 \\ 0 & \frac{2T}{\sigma_\varepsilon^4} \end{pmatrix} = \begin{pmatrix} \mathcal{I}_{\tau\tau} & 0 \\ 0 & \mathcal{I}_{\varepsilon\varepsilon} \end{pmatrix}. \quad (\text{A-47})$$

This expression follows by direct calculation. Asymptotic normality of both  $\hat{\tau}_x$  and  $\hat{\sigma}_\varepsilon^2$  follows by the usual arguments. We can determine the asymptotic variance of  $\widehat{\langle X, X \rangle}^{(w)}$  via

$$\begin{aligned} \text{var} \left\{ \widehat{\langle X, X \rangle}^{(w)} \right\} &= T^2 \text{var} \{ \hat{\tau}_x \} \\ &= T \frac{\sigma_\varepsilon}{\tau_X^{1/2}} 16 \tau_X^2 \sqrt{\Delta t}. \end{aligned} \quad (\text{A-48})$$

We see that the variance depends on the length of the time course, the inverse of the signal to noise ratio, the square root of the sampling period and the fourth power of the “average standard deviation” of the  $X_t$  process.

## B. PROOF OF THEOREM 2

We now wish to use these results to deduce properties of  $\hat{\sigma}$ . Firstly using the well known *invariance* of maximum likelihood estimators to transfer the estimators of  $\bar{\sigma}_X^2$  and  $\sigma_\varepsilon^2$  to estimators of  $\langle X, X \rangle_T$ . We therefore take

$$\widehat{\langle X, X \rangle}_T^{(m_1)} = \sum_{k=0}^{N-1} \hat{\mathcal{S}}_{kk}^{(X)}(\hat{L}_k) = \sum_{k=0}^{N-1} \hat{L}_k \hat{\mathcal{S}}_{kk}^{(Y)}$$

It therefore follows that with  $\hat{\tau}_X = \tau_X + \delta\tau_X$  and  $\hat{\sigma}_\varepsilon^2 = \sigma_\varepsilon^2 + \delta\sigma_\varepsilon^2$

$$\begin{aligned} \mathbb{E} \left\{ \widehat{\langle X, X \rangle}_T^{(m_1)} \right\} &= \sum_{k=0}^{N-1} \mathbb{E} \left\{ \left( \frac{\bar{\sigma}_X^2 + \delta\sigma_X^2}{\bar{\sigma}_X^2 + \delta\sigma_X^2 + (\sigma_\varepsilon^2 + \delta\sigma_\varepsilon^2) |2 \sin(\pi f_k \Delta t)|^2} \right) \hat{\mathcal{S}}_{kk}^{(Y)} \right\} \\ &= \sum_{k=0}^{N-1} \mathbb{E} \left\{ \left( \frac{\bar{\sigma}_X^2 + \delta\sigma_X^2}{1 + \frac{[\delta\sigma_X^2 + \delta\sigma_\varepsilon^2] |2 \sin(\pi f_k \Delta t)|^2}{\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2}} \right) \frac{\hat{\mathcal{S}}_{kk}^{(Y)}}{\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2} \right\} \\ &= \sum_{k=0}^{N-1} \mathbb{E} \left\{ (\bar{\sigma}_X^2 + \delta\sigma_X^2) \left( 1 - \frac{[\delta\sigma_X^2 + \delta\sigma_\varepsilon^2] |2 \sin(\pi f_k \Delta t)|^2}{(\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2)} \right) \frac{\hat{\mathcal{S}}_{kk}^{(Y)}}{\bar{\sigma}_X^2 + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2} \right\} \\ &= \sum_{k=0}^{N-1} [\bar{\sigma}_X^2 + O(\Delta t^{5/4})] + O(\sqrt{\Delta t} \log(\Delta t)) = \mathbb{E} \{ \langle X, X \rangle_T \} + O(\sqrt{\Delta t} \log(\Delta t)) \\ &\quad + O(\sqrt[4]{\Delta t}). \end{aligned}$$



This implies that the estimator is asymptotically unbiased. We can also note that the variance of the new estimator is given by:

$$\begin{aligned}
\text{var} \left\{ \widehat{\langle X, X \rangle}_T^{(m_1)} \right\} &= \sum_j \sum_k \text{cov} \{ \widehat{L}_j \widehat{\mathcal{S}}_{jj}^{(Y)}, \widehat{L}_k \widehat{\mathcal{S}}_{kk}^{(Y)} \} \\
&= \sum_j \sum_k \text{cov} \left\{ \frac{\widehat{L}_j}{L_j} L_j \widehat{\mathcal{S}}_{jj}^{(Y)}, \frac{\widehat{L}_k}{L_k} L_k \widehat{\mathcal{S}}_{kk}^{(Y)} \right\} \\
&= \sum_j \sum_k \text{cov} \left\{ \left( 1 + \frac{\delta\tau_X}{\tau_X} - \frac{\delta\tau_X \Delta t + \delta\sigma_\varepsilon^2 |2 \sin(\pi f_j \Delta t)|^2}{\tau_X \Delta t + \sigma_\varepsilon^2 |2 \sin(\pi f_j \Delta t)|^2} \right) L_j \widehat{\mathcal{S}}_{jj}^{(Y)}, \right. \\
&\quad \left. \left( 1 + \frac{\delta\tau_X}{\tau_X} - \frac{\delta\tau_X \Delta t + \delta\sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2}{\tau_X \Delta t + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2} \right) L_k \widehat{\mathcal{S}}_{kk}^{(Y)} \right\}.
\end{aligned}$$

Then

$$\begin{aligned}
\text{var} \left\{ \widehat{\langle X, X \rangle}_T^{(m_1)} \right\} &= \sum_j \sum_k \left\{ \text{cov} \{ L_j \widehat{\mathcal{S}}_{jj}^{(Y)}, L_k \widehat{\mathcal{S}}_{kk}^{(Y)} \} + \text{cov} \left\{ \frac{\delta\tau_X}{\tau_X} L_j \widehat{\mathcal{S}}_{jj}^{(Y)}, L_k \widehat{\mathcal{S}}_{kk}^{(Y)} \right\} \right. \\
&\quad + \text{cov} \left\{ L_j \widehat{\mathcal{S}}_{jj}^{(Y)}, \frac{\delta\tau_X}{\tau_X} L_k \widehat{\mathcal{S}}_{kk}^{(Y)} \right\} \\
&\quad - \text{cov} \left\{ \frac{\delta\tau_X \Delta t + \delta\sigma_\varepsilon^2 |2 \sin(\pi f_j \Delta t)|^2}{\tau_X \Delta t + \sigma_\varepsilon^2 |2 \sin(\pi f_j \Delta t)|^2} L_j \widehat{\mathcal{S}}_{jj}^{(Y)}, L_k \widehat{\mathcal{S}}_{kk}^{(Y)} \right\} \\
&\quad - \text{cov} \left\{ L_j \widehat{\mathcal{S}}_{jj}^{(Y)}, \frac{\delta\tau_X \Delta t + \delta\sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2}{\tau_X \Delta t + \sigma_\varepsilon^2 |2 \sin(\pi f_k \Delta t)|^2} L_k \widehat{\mathcal{S}}_{kk}^{(Y)} \right\} \\
&\quad \left. + \text{cov} \left\{ \frac{\delta\tau_X}{\tau_X} L_j \widehat{\mathcal{S}}_{jj}^{(Y)}, \frac{\delta\tau_X}{\tau_X} L_k \widehat{\mathcal{S}}_{kk}^{(Y)} \right\} + \dots \right\} \\
&= \sum_j \sum_k \left\{ \delta_{jk} \sigma_X^4 + L_j L_k \text{cov} \left\{ \frac{\delta\tau_X}{\tau_X} \widehat{\mathcal{S}}_{jj}^{(Y)}, \widehat{\mathcal{S}}_{kk}^{(Y)} \right\} + \dots \right\}
\end{aligned}$$

By looking at the individual terms of this expression, and noting that the estimated renormalized variance  $\widehat{\tau}_X = \tau_X + \delta\tau_X$  and  $\widehat{\sigma}_\varepsilon^2 = \sigma_\varepsilon^2 + \delta\sigma_\varepsilon^2$  are linear combinations of  $\widehat{\mathcal{S}}_{kk}^{(Y)}$ , we can deduce the stated order terms, by again noting the  $\sqrt{\Delta t}$  order of the important terms. However to leading order, this estimator will perform identically to  $\widehat{\langle X, X \rangle}_T^{(w)}$  in terms of variance.

### C. PROOF OF TIME DOMAIN FORM

The integral can be calculated from first principles using complex-variables with  $z = e^{2i\pi f}$ . Thus  $dz/df = 2i\pi z$  or  $df = dz/(2i\pi z)$ . (28) takes the form

$$\ell_\tau = \frac{1}{2i\pi} \oint_{|z|=1} \frac{\overline{\sigma}_X^2}{\overline{\sigma}_X^2 z - \sigma_\varepsilon^2 [z - 1]^2} z^\tau dz. \tag{C-49}$$

We need the poles, or:

$$\overline{\sigma}_X^2 z - \sigma_\varepsilon^2 [z - 1]^2 = 0 \iff z = 1 + \frac{\overline{\sigma}_X^2}{2\sigma_\varepsilon^2} \pm \sqrt{\frac{\overline{\sigma}_X^2}{\sigma_\varepsilon^2} + \frac{\overline{\sigma}_X^4}{4\sigma_\varepsilon^4}} = z^\pm$$

If  $\left| \frac{\sigma_\varepsilon^2}{\sigma_X^2} \right| < 1$  we have

$$\begin{aligned}
 z^- &= 1 + \frac{\bar{\sigma}_X^2}{2\sigma_\varepsilon^2} - \frac{\bar{\sigma}_X^2}{2\sigma_\varepsilon^2} \sqrt{1 + \frac{4\sigma_\varepsilon^2}{\sigma_X^2}} \\
 &= 1 + \frac{\bar{\sigma}_X^2}{2\sigma_\varepsilon^2} - \frac{\bar{\sigma}_X^2}{2\sigma_\varepsilon^2} \left( 1 + \frac{1}{2} \frac{4\sigma_\varepsilon^2}{\sigma_X^2} + \frac{1}{4} \frac{(-1)}{2} \left[ \frac{4\sigma_\varepsilon^2}{\sigma_X^2} \right]^2 + \mathcal{O} \left( \frac{\sigma_\varepsilon^6}{\sigma_X^6} \right) \right) \\
 &= \frac{\sigma_\varepsilon^2}{\sigma_X^2} + \mathcal{O} \left( \frac{\sigma_\varepsilon^4}{\sigma_X^4} \right) \\
 z^+ &= \frac{\sigma_X^2}{\sigma_\varepsilon^2} + \dots
 \end{aligned}$$

We then note that:

$$\begin{aligned}
 \ell_\tau &= -\frac{1}{2i\pi} \oint_{|z|=1} \frac{\bar{\sigma}_X^2/(\sigma_\varepsilon^2)}{-\bar{\sigma}_X^2/(\sigma_\varepsilon^2)z + [z-1]^2} z^\tau dz = -\frac{1}{2i\pi} \oint_{|z|=1} \frac{\bar{\sigma}_X^2/(\sigma_\varepsilon^2)}{(z-z^-)(z-z^+)} z^\tau dz \\
 &= \frac{2i\pi}{2i\pi} \frac{\bar{\sigma}_X^2/(\sigma_\varepsilon^2)}{z^+ - \frac{\sigma_\varepsilon^2}{\sigma_X^2} + \mathcal{O} \left( \frac{\sigma_\varepsilon^4}{\sigma_X^4} \right)} \left( \frac{\sigma_\varepsilon^2}{\sigma_X^2} \right)^\tau = \left( \frac{\sigma_\varepsilon^2}{\sigma_X^2} \right)^\tau + \mathcal{O} \left( \frac{\sigma_\varepsilon^{2\tau+2}}{\sigma_X^{2\tau+2}} \right)
 \end{aligned}$$

If on the other hand you consider  $\left| \frac{\sigma_\varepsilon^2}{\sigma_X^2} \right| > 1$  which in many scenarios is more realistic then we find that:

$$\begin{aligned}
 z^- &= 1 + \frac{\bar{\sigma}_X^2}{2\sigma_\varepsilon^2} - \frac{\bar{\sigma}_X}{\sigma_\varepsilon} \sqrt{1 + \frac{\sigma_X^2}{4\Delta t \sigma_\varepsilon^2}} = 1 + \frac{\bar{\sigma}_X^2}{2\sigma_\varepsilon^2} - \frac{\bar{\sigma}_X}{\sigma_\varepsilon} \left( 1 + \frac{1}{2} \frac{\sigma_X^2}{4\sigma_\varepsilon^2} \right) \\
 &= 1 - \frac{\bar{\sigma}_X}{\sigma_\varepsilon} + \mathcal{O} \left( \frac{\bar{\sigma}_X^2}{\sigma_\varepsilon^2} \right) \\
 z^+ &= 1 + \frac{\bar{\sigma}_X}{\sigma_\varepsilon}
 \end{aligned}$$

In this case we find that

$$\ell_\tau = \frac{\bar{\sigma}_X^2/(\sigma_\varepsilon^2)}{2\frac{\bar{\sigma}_X}{\sigma_\varepsilon} + \mathcal{O} \left( \frac{\bar{\sigma}_X^2}{\sigma_\varepsilon^2} \right)} \frac{[1 - \frac{\bar{\sigma}_X}{\sigma_\varepsilon}]^\tau}{\sigma_\varepsilon^2} = \frac{\bar{\sigma}_X}{2\sigma_\varepsilon} \left( 1 - \frac{\bar{\sigma}_X}{\sigma_\varepsilon} \right)^\tau + \mathcal{O} \left( \frac{\bar{\sigma}_X^2}{2\sigma_\varepsilon^2} \left( 1 - \frac{\bar{\sigma}_X}{\sigma_\varepsilon} \right)^\tau \right).$$

In both cases the decay of the filter is geometric. We note that in most practical examples  $L_k$  decays very rapidly in  $k$ . Therefore, we do not need to integrate between  $-1/2$  to  $1/2$ , and only need to integrate over  $-1/\pi$  to  $1/\pi$ . In this range of  $f$  we find that for smallish remainder term  $R_3$  we have:  $\sin^2(\pi f) = \pi^2 f^2 + R_3(f\pi)$ . Then we note

$$\begin{aligned}
 \ell_\tau &= \int_{-\frac{1}{\pi}}^{\frac{1}{\pi}} \frac{\bar{\sigma}_X^2}{\bar{\sigma}_X^2 + 4\sigma_\varepsilon^2 \pi^2 f^2 + R_3(f\pi)} e^{2i\pi f \tau} df + C \\
 &= \frac{\bar{\sigma}_X}{2\sigma_\varepsilon} \int_{-\infty}^{\infty} \left[ 2 \frac{\bar{\sigma}_X/(\sigma_\varepsilon)}{\frac{\sigma_X^2}{\sigma_\varepsilon^2} + 4\pi^2 f^2} + R_4(f) \right] e^{2i\pi f \tau} df + C \\
 &= \frac{\bar{\sigma}_X}{2\sigma_\varepsilon} e^{-\frac{\bar{\sigma}_X |\tau|}{\sigma_\varepsilon}} + C.
 \end{aligned}$$

Thus we are smoothing the autocovariance sequence with a smoothing window that becomes a delta function as  $\bar{\sigma}_X/\sigma_\varepsilon \rightarrow \infty$ . It is reasonable that this non-dimensional quantity arises as an important factor.